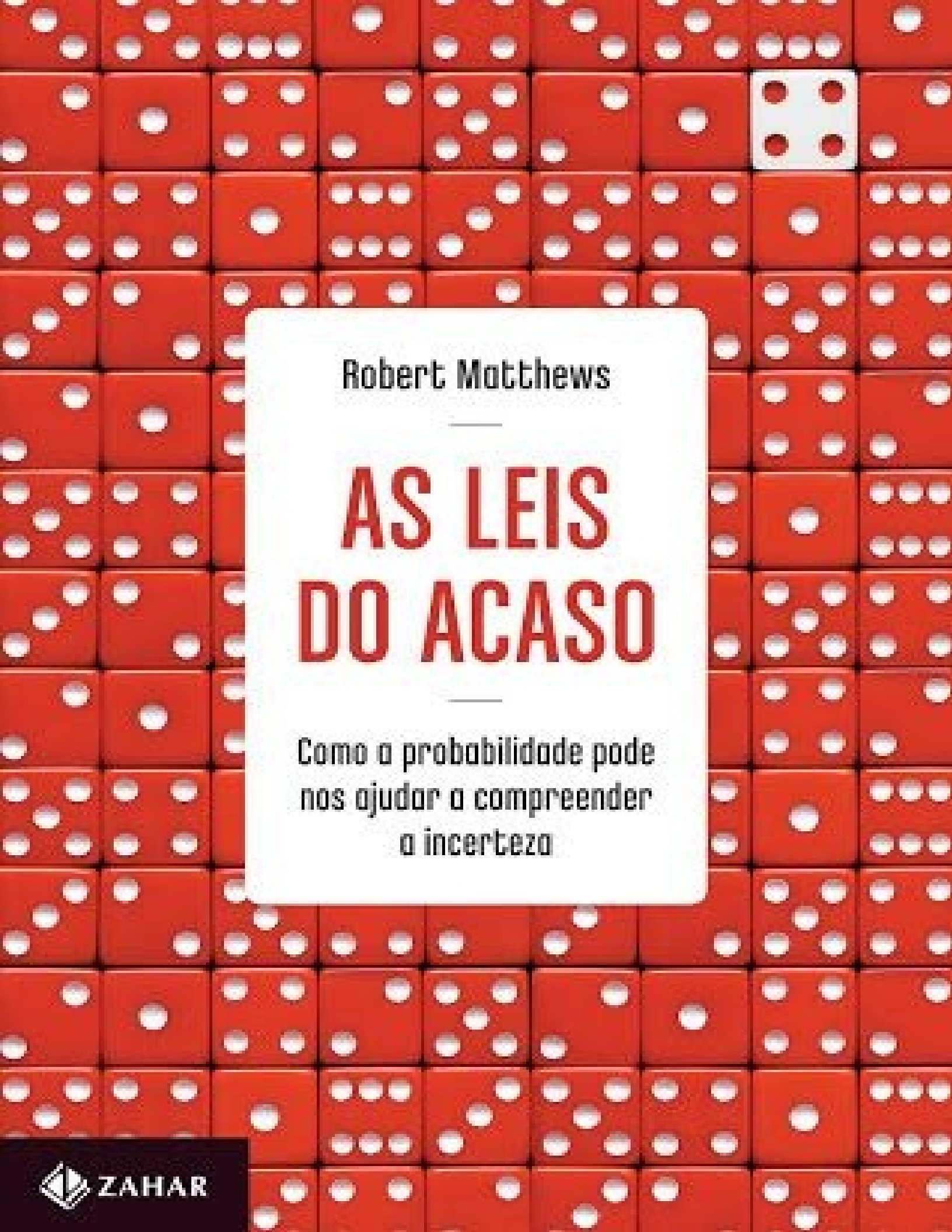




Robert Matthews

AS LEIS DO ACASO

Como a probabilidade pode
nos ajudar a compreender
a incerteza



Encontre mais livros como este no [e-Livros](#)

[e-Livros.xyz](#)

[e-Livros.site](#)

[e-Livros.website](#)



TRADIÇÃO EM
COMPARTILHAR
CONHECIMENTO

Robert Matthews

As leis do acaso

Como a probabilidade pode nos ajudar a compreender a
incerteza

Tradução:

George Schlesinger

Revisão técnica:

Samuel Jurkiewicz

professor da Politécnica e da Coppe/UFRJ



*Para Denise,
A pessoa mais esperta que conheço,
e que, imprevisivelmente, apostou suas cartas em mim.*

Sumário

Introdução

1. O lançador de moedas prisioneiro dos nazistas
2. O que *realmente* significa a lei das médias
3. O obscuro segredo do teorema áureo
4. A primeira lei da ausência de leis
5. Quais são as chances *disso*?
6. Pensar de modo independente não inclui gema de ovo
7. Lições aleatórias da loteria
8. Aviso: há muito X por aí
9. Por que o espetacular tantas vezes vira “mais ou menos”
10. Se você não sabe, vá pelo aleatório
11. Nem sempre é ético fazer a coisa certa
12. Como uma “boi-bagem” deflagrou uma revolução
13. Como vencer os cassinos no jogo deles
14. Onde os espertinhos se dão mal
15. A regra áurea das apostas
16. Garantir – ou arriscar?
17. Fazer apostas melhores no cassino da vida
18. Diga a verdade, doutor, quais as minhas chances?

19. Isso não é uma simulação! Repito, isso não é uma simulação!
20. A fórmula milagrosa do reverendo Bayes
21. O encontro do dr. Turing com o reverendo Bayes
22. Usando Bayes para julgar melhor
23. Um escândalo de significância
24. Esquivando-se da espantosa máquina de bobagens
25. Use aquilo que você já sabe
26. Desculpe, professor, mas não engulo essa
27. A assombrosa curva para tudo
28. Os perigos de pensar que tudo é normal
29. Irmãs feias e gêmeas malvadas
30. Até o extremo
31. Assista a um filme de Nicolas Cage e morra
32. Temos de traçar a linha em algum lugar
33. Jogar com os mercados *não* é uma ciência precisa
34. Cuidado com geeks criando modelos

Notas

Agradecimentos

Índice remissivo

Introdução

NUMA TARDE DE DOMINGO de abril de 2004, um inglês de 32 anos entrou no Plaza Hotel & Casino, em Las Vegas, com todas as suas posses mundanas. Elas consistiam em uma muda de roupa de baixo e um cheque. Ashley Revell tinha vendido tudo que possuía para levantar a quantia de US\$ 135 300, impressa no cheque; até o smoking que ele vestia era alugado. Depois de trocar o cheque por uma pilha de fichas desoladoramente pequena, Revell dirigiu-se à roleta e fez uma coisa extraordinária. Apostou tudo num só resultado: quando a bolinha branca parasse, ela cairia no vermelho.

A decisão de Revell de escolher essa cor pode ter sido impulsiva, mas o fato em si não foi. Ele planejara aquilo durante meses. Conversara sobre o assunto com amigos, que acharam a ideia brilhante, e com a sua família, que achou-a péssima. Os cassinos tampouco aprovaram; talvez tivessem medo de entrar para o folclore de Las Vegas como “o cassino em que um homem apostou tudo e perdeu”. Decerto o gerente do Plaza tinha um ar solene quando Revell colocou as fichas sobre a mesa, e lhe perguntou se tinha certeza de que queria ir em frente. Mas nada parecia capaz de deter Revell. Cercado por um grande grupo de espectadores, ele esperou ansiosamente o crupiê jogar a bolinha na roleta. Então, num gesto único e rápido, deu um passo adiante e pôs todas as fichas no vermelho. Assistiu à bolinha diminuir de velocidade, percorrer a trajetória em espirais, ricocheteando em várias casas, e finalmente parar... na casa número 7. Vermelho.

Naquele momento Revell dobrou seu patrimônio líquido para US\$ 270 600. A multidão o ovacionou e seus amigos o abraçaram – e seu pai pesarosamente o chamou de “menino malcriado”. É improvável que a maioria das pessoas adotasse visão mais severa acerca das ações de Revell naquele dia; na melhor das hipóteses, o julgariam mal aconselhado, sem dúvida alguma insensato e

possivelmente insano. Pois decerto nem os bilionários, para quem essas quantias são troco miúdo, teriam jogado a bolada toda de uma vez. Qualquer ser racional não teria dividido a quantia em apostas menores, para ao menos conferir se dona Sorte estava por perto?

Mas aí está o lance: uma vez decidido, Revell fez a coisa certa. As leis da probabilidade mostram que não há meio mais seguro de dobrar o patrimônio num cassino que fazer o que ele fez, e apostar tudo num só giro da roleta. Sim, o jogo é injusto: as chances da roleta são deliberadamente – e legalmente – contra você. Sim, havia mais de 50% de chance de perder tudo. No entanto, por mais bizarro que possa parecer, nessas situações, a melhor estratégia é apostar grande e com audácia. Qualquer coisa mais tímida reduz as chances de sucesso. O próprio Revell provara isso durante os preparativos para a grande aposta. Nos dias anteriores apostara vários milhares de dólares no cassino, e tudo que conseguira foi perder US\$ 1 mil. Sua maior esperança de duplicar seu dinheiro residia em trocar o “senso comum” pelos ditames das leis da probabilidade.

Então, devemos todos seguir o exemplo de Revell, vender tudo que possuímos e nos dirigir ao cassino mais próximo? Claro que não; existem maneiras muito melhores, embora mais chatas, de tentar duplicar seu dinheiro. Todavia, uma coisa é certa: todas elas envolvem probabilidade em uma de suas muitas roupagens: como chance, risco ou grau de crença.

Todos nós sabemos que há poucas certezas na vida, exceto a morte e os impostos. Mas poucos de nós se sentem à vontade na presença da probabilidade. Ela ameaça qualquer sensação que tenhamos de controlar os fatos, sugerindo que todos poderíamos nos tornar o que Shakespeare chamou de “bobo da Fortuna”. Ela tem levado alguns a acreditar em deuses volúveis, outros a negar sua supremacia. Einstein recusava-se a acreditar que Deus joga dados com o Universo. No entanto, a própria ideia de dar sentido à probabilidade parece contraditória: o acaso, por definição, não está para além da compreensão? Essa lógica pode ressaltar um dos grandes mistérios da história intelectual. Por que, apesar de sua óbvia utilidade, demorou tanto tempo para surgir uma teoria confiável da probabilidade? Ainda que houvesse jogos de azar no Egito Antigo,

há mais de 5 500 anos, foi só no século XVII que alguns pensadores ousados desafiaram seriamente a visão sintetizada por Aristóteles, de que “não pode haver conhecimento demonstrativo da probabilidade”.

Não adianta nada o fato de a probabilidade desafiar com tanta frequência nossas intuições. Pensemos nas coincidências: em termos gerais, quais são as chances de, num jogo de futebol, haver dois jogadores que façam aniversário em dias consecutivos? Como há 365 dias no ano, e 22 jogadores, alguém pode dizer que a chance é menor que uma em dez. Na verdade, as leis da probabilidade revelam que a verdadeira resposta é mais ou menos 90%. Você não acredita? Então confira os aniversários dos jogadores de algumas partidas, e veja você mesmo. Mesmo assim, é difícil não pensar que está acontecendo algo muito estranho. Afinal, se estiver entre um grupo de tamanho semelhante e perguntar se alguém nasceu *no mesmo dia que você*, é muito pouco provável que encontre alguém. Até problemas simples, de lançamento de moedas e dados, parecem desafiar o senso comum. Com uma moeda honesta, certamente obter cara em vários lançamentos seguidos torna coroa mais provável, certo? Se você está batalhando para ver por que isso não é verdade, não se preocupe: um dos grandes matemáticos do Iluminismo jamais conseguiu captar isso.

Um dos objetivos deste livro é mostrar como compreender essas manifestações cotidianas da probabilidade revelando suas leis subjacentes e como aplicá-las. Veremos como usar essas leis para *predizer* coincidências, tomar decisões melhores nos negócios e na vida, e dar sentido a tudo, de diagnósticos médicos a conselhos de investimentos.

Mas este não é só um livro sobre boas dicas e sugestões convenientes. Meu principal objetivo é mostrar como as leis da probabilidade são capazes de muita coisa além de apenas entender os eventos probabilísticos. Elas são também a arma preferida para qualquer pessoa que tenha necessidade de transformar evidência em sacação. Desde a identificação dos riscos para a saúde e das novas drogas para lidar com eles até progressos na nossa compreensão do cosmo, as leis da probabilidade têm se mostrado cruciais para separar impurezas aleatórias do ouro das evidências.

Agora outra revolução está em andamento, uma revolução centrada nas próprias leis da probabilidade. Hoje fica mais evidente que, na busca do conhecimento, essas leis são bem mais poderosas do que se pensava. Mas ter acesso a esse poder exige uma reinterpretação radical da probabilidade – o que até há pouco provocava amargas discussões. A controvérsia que durou décadas hoje some diante da evidência de que os chamados métodos bayesianos podem transformar a ciência, a tecnologia e a medicina. Até aqui, muito pouco disso tem chegado ao público. Neste livro, eu conto a história, em geral espantosa, do surgimento dessas técnicas, as polêmicas que elas provocaram e como todos nós podemos usá-las para dar sentido a tudo, desde a previsão do tempo até a credibilidade de novos argumentos científicos.

Qualquer pessoa que queira dominar as leis da probabilidade, porém, deve saber quais são as limitações dessas leis e quando se faz delas um uso impróprio. Agora está ficando claro que os métodos que constam dos livros-texto, e nos quais os pesquisadores se apoiaram durante muito tempo para tirar conclusões a partir dos dados, na maioria das vezes estão forçados para além de seus limites próprios. Avisos sobre as possíveis consequências catastróficas dessa prática vêm circulando nos meios acadêmicos durante décadas. Mais uma vez, muito pouco desse escândalo emergente chega ao domínio público. Este livro busca remediar o problema. Ao fazê-lo, ele recorre às minhas próprias contribuições para a bibliografia de pesquisa e inclui formas de identificar quando a evidência e os métodos aplicados são forçados demais.

A necessidade de compreender probabilidade, risco e incerteza nunca foi mais urgente. Em face de agitações políticas, tumultos nos mercados financeiros e uma interminável ladainha sobre riscos, ameaças e calamidades, todos nós ficamos ansiosos por uma certeza. Na verdade, ela nunca existiu. Mas isso não é razão para fatalismos – ou para a recusa em aceitar a realidade.

A mensagem central deste livro é que, apesar de não podermos nos livrar da probabilidade, do risco e da incerteza, agora temos as ferramentas para adotá-los e vencer.

1. O lançador de moedas prisioneiro dos nazistas

NA PRIMAVERA DE 1940, John Kerrich saiu de casa para visitar os parentes da esposa – o que não era pouca coisa, porque Kerrich morava na África do Sul e os parentes estavam na Dinamarca, a 12 mil quilômetros de distância. E no momento em que chegou a Copenhague deve ter desejado ter ficado em casa. Apenas alguns dias antes, a Dinamarca fora invadida pela Alemanha nazista. Milhares de soldados avançaram como formigas sobre a fronteira, numa arrasadora demonstração de Blitzkrieg. Em poucas horas os nazistas tinham vencido a resistência e assumido o controle. Durante as semanas que se seguiram, dedicaram-se a prender estrangeiros inimigos e levá-los para campos de concentração. Logo Kerrich se viu entre eles.

Poderia ter sido pior. Ele foi para um campo na Jutlândia, dirigido pelo governo dinamarquês e, conforme relatou depois, administrado de “forma realmente admirável”.¹ Mesmo assim, sabia que enfrentaria muitos meses, possivelmente anos, sem qualquer estímulo intelectual – o que não era uma perspectiva feliz para um professor de matemática da Universidade de Witwatersrand. Circulando pelo campo em busca de algo para ocupar seu tempo, teve a ideia de um projeto matemático que exigia equipamento mínimo, mas que poderia ser instrutivo para os outros. Decidiu embarcar num estudo abrangente sobre o funcionamento da probabilidade na mais básica de suas manifestações: o resultado do lançamento de uma moeda.

Kerrich já tinha familiaridade com a teoria desenvolvida pelos matemáticos para compreender o funcionamento da probabilidade. Agora, percebeu ele, tinha a rara ocasião de testar essa teoria com uma porção de dados simples, da vida real. Então, uma vez terminada a guerra – presumindo, claro, que sobrevivesse a ela –, seria capaz de voltar à universidade equipado não só com os fundamentos teóricos das leis da probabilidade, mas também com evidências sólidas para que

elas ganhassem confiança. E isso seria inestimável para explicar a seus alunos as previsões, evidentemente contrárias ao senso comum, das leis da probabilidade.

Ele queria que seu estudo fosse o mais abrangente e confiável possível, e isso significava lançar uma moeda e registrar o resultado pelo máximo tempo que pudesse aguentar. Felizmente, encontrou alguém disposto a compartilhar o tédio, um colega prisioneiro chamado Eric Christensen. E assim, juntos, montaram uma mesa, estenderam um pano por cima e, com um movimento do dedão, lançaram uma moeda cerca de trinta centímetros de altura.

Para que fique registrado, o lançamento deu coroa.

Muita gente provavelmente acha que pode adivinhar como as coisas aconteceram a partir daí. À medida que o número de lançamentos aumentasse, a conhecida lei das médias iria garantir que começariam a se equilibrar as vezes em que sairia cara ou coroa. De fato, Kerrich descobriu que, por volta do centésimo lançamento, os números de caras e de coroas eram bastante semelhantes: 44 caras contra 56 coroas.

Mas aí começou a acontecer uma coisa estranha. À medida que as horas e os lançamentos avançavam, as caras começaram a ultrapassar as coroas. Por volta do lançamento 2 mil, a diferença tinha mais que duplicado, e as caras tinham uma dianteira de 26 sobre as coroas. Na altura do 4 mil, a diferença chegava a 58. A discrepância parecia se tornar maior.

No momento em que Kerrich fez uma pausa – no lançamento 10 mil –, a moeda tinha dado cara 5 067 vezes, excedendo o número de coroas pela robusta margem de 134. Longe de desaparecer, a discrepância entre caras e coroas continuara a aumentar. Haveria algo de errado com o experimento? Ou teria Kerrich descoberto uma falha na lei das médias? Kerrich e Christensen tinham feito o melhor para excluir lançamentos duvidosos, e, quando fecharam os números, viram que a lei das médias não fora em absoluto violada. O problema real não era com a moeda nem com a lei, mas com a visão comumente adotada acerca do que diz a lei. O experimento simples de Kerrich tinha na verdade feito

o que ele queria fazer. Demonstrara uma das grandes concepções errôneas sobre o funcionamento da probabilidade.

Indagadas sobre o que diz a lei das médias, muitas pessoas falam algo do tipo: “A longo prazo, tudo se equilibra.” Como tal, a lei é uma fonte de consolo quando temos uma sequência de azar, ou quando os nossos inimigos parecem estar em ascensão. Torcedores no esporte muitas vezes invocam isso quando se sentem vítimas de um cara ou coroa perdido ou da má decisão de uma arbitragem. Ganhar algumas, perder outras... no fim tudo se equilibra.

Bem, sim e não. Sim, de fato há uma lei das médias em ação no nosso Universo. Sua existência não foi apenas demonstrada experimentalmente, mas foi provada do ponto de vista matemático. Ela se aplica não só ao nosso Universo, mas em todo Universo com as mesmas regras matemáticas que o nosso; nem as leis da física podem reivindicar isso. Mas não, a lei não implica que “no fim tudo se equilibra”. Como veremos em outros capítulos, definir o que ela significa com precisão exigiu um volume imenso de esforços de alguns dos maiores matemáticos do último milênio. Eles ainda discutem sobre a lei, mesmo agora. Sabe-se que com frequência os matemáticos exigem um nível de exatidão que o resto de nós consideraria ridiculamente pedante. Mas nesse caso eles estão certos em serem exigentes. Pois acontece que saber o que diz a lei das médias com precisão é uma das chaves para compreender como a probabilidade funciona no nosso mundo – e como usar essa compreensão em nosso proveito. A chave para essa compreensão reside em estabelecer exatamente a que nos referimos por “no fim tudo se equilibra”. Em particular, o que é esse “tudo”?

Isso soa perigosamente parecido com um exercício filosófico de olhar para o próprio umbigo, mas o experimento de Kerrich aponta para a resposta certa. Muita gente acha que esse “tudo” onde os eventos se equilibram a longo prazo são os números absolutos de caras e coroas.

Então, por que a moeda gerou um resultado muito maior de uma face que de outra? A resposta curta é: porque era a probabilidade cega, aleatória, que atuava em cada lançamento da moeda, tornando ainda mais improvável a coincidência exata dos números absolutos de caras e coroas. O que aconteceu com a lei das

médias? Ela está viva e passa bem, o caso é que simplesmente não se aplica aos números absolutos de caras e coroas. É bastante óbvio que não podemos dizer com toda a certeza como irão se comportar eventos aleatórios individuais. Mas podemos dizer algo sobre eles se descermos para um nível de conhecimento ligeiramente inferior – e perguntarmos como os eventos aleatórios se comportam *em média*.

No caso do lançamento de uma moeda, não podemos afirmar com certeza quando teremos “cara” ou “coroa”, nem quantas vezes irá sair cada face. Mas, considerando que há apenas dois resultados, e que eles são igualmente prováveis, podemos dizer que devem aparecer com igual *frequência* – ou seja, 50% das vezes.

Isso, por sua vez, mostra exatamente o que é esse “tudo” que “equilibra os eventos a longo prazo”. Não são os *números absolutos* de caras e coroas, sobre os quais não podemos afirmar nada com certeza. São suas *frequências relativas*: o número de vezes que cada um aparece, como proporção do número total de oportunidades que nós lhe damos de aparecer.

Essa é a verdadeira lei das médias, e foi o que Kerrich e Christensen viram em seu experimento. À medida que os lançamentos se acumulavam, as frequências relativas de caras e coroas – isto é, sua quantidade dividida pela quantidade total de lançamentos – foram chegando cada vez mais perto. Quando o experimento terminou, essas frequências tinham uma margem de 1% de serem idênticas (50,67% de caras contra 49,33% de coroas). Em agudo contraste, os números absolutos de caras e coroas iam se afastando mais e mais (ver Tabela).

A lei das médias nos diz que, se quisermos entender a ação do acaso sobre os eventos, devemos focalizar não cada evento individual, mas suas frequências relativas. Sua importância se reflete no fato de que muitas vezes elas são consideradas a medida da característica mais básica de todos os eventos aleatórios: sua *probabilidade*.

Nº DE LANÇAMENTOS	Nº DE CARAS	Nº DE COROAS	DIFERENÇA (CARAS – COROAS)	FREQUÊNCIA DE CARAS
----------------------	-------------	--------------	-------------------------------	------------------------

10	4	6	-2	40,00%
100	44	56	-12	44,00%
500	255	245	+10	51,00%
1 000	502	498	+4	50,20%
5 000	2 533	2 467	+66	50,66%
10 000	5 067	4 933	+134	50,67%

A verdadeira lei das médias e o que realmente significa “no final tudo se equilibra”.

UM LANÇAMENTO DE MOEDA É REALMENTE JUSTO?

Em geral, considera-se aleatório o lançamento de moeda, mas pode-se prever como ela cai – pelo menos em teoria. Em 2008, uma equipe da Universidade Técnica de Łódź, na Polônia,² analisou a mecânica de uma moeda de verdade caindo sob a ação da resistência do ar. A teoria é muito complexa, mas revelou que o comportamento da moeda é previsível até atingir o solo. Então se instala o comportamento “caótico”, com pequenas diferenças produzindo resultados radicalmente diferentes. Isso, por sua vez, sugeriu que lançamentos de moedas apanhadas em pleno ar podem ter um ligeiro viés. Essa possibilidade também foi investigada por uma equipe orientada pelo matemático Persi Diaconis, da Universidade Stanford.³ Eles descobriram que moedas apanhadas no ar têm uma leve tendência a acabar no mesmo estado em que começaram. O viés, porém, é incrivelmente pequeno. Assim, os resultados de se lançar uma moeda podem de fato ser considerados aleatórios, quer ela seja apanhada no ar, quer caia no chão.

Assim, por exemplo, se rolarmos um dado mil vezes, a chance aleatória tem muito pouca probabilidade de fazer com que os números de 1 a 6 apareçam precisamente a mesma quantidade de vezes; essa é uma afirmativa acerca de resultados individuais, sobre os quais não se pode dizer nada com certeza. Graças à lei das médias, porém, podemos esperar que as *frequências relativas* dos diferentes resultados sejam em torno de $\frac{1}{6}$ do total dos lances dos dados – e cheguem ainda mais perto dessa proporção exata quanto mais vezes o dado for rolado. Essa proporção exata é o que chamamos de probabilidade de cada número aparecer (embora, como veremos adiante, não seja o único modo de

pensar a probabilidade). Para algumas coisas – como a moeda, o dado ou o baralho – podemos ter uma noção da probabilidade a partir das propriedades fundamentais que governam os vários resultados (o número de lados, os naipes das cartas etc.) Assim, é possível dizer que, a longo prazo, as frequências relativas dos resultados devem se aproximar cada vez mais dessa probabilidade. Se isso não acontecer, devemos começar a nos perguntar por que nossas crenças se mostraram mal fundamentadas.

Conclusão

A lei das médias nos diz que, quando sabemos – ou desconfiamos – que estamos lidando com eventos envolvendo um elemento de acaso, devemos focalizar não os eventos em si, mas sua frequência relativa – isto é, o número de vezes que cada evento ocorre em proporção ao número total de oportunidades.

2. O que *realmente* significa a lei das médias

A LEI DAS MÉDIAS nos avisa que, ao lidar com eventos aleatórios, são suas frequências relativas, e não os números brutos, que devemos focalizar. Mas se você está lutando para abandonar a ideia de que os números brutos “se equilibram a longo prazo”, não se atormente; você está em boa companhia. Jean-Baptiste le Rond d’Alembert, um dos grandes matemáticos do Iluminismo, estava seguro de que uma sequência de caras ao lançar uma moeda tornava coroa cada vez mais provável.

Mesmo hoje, muitas pessoas geralmente experientes jogam fora um bom dinheiro em cassinos e casas de aposta acreditando que uma sequência de azar torna a boa sorte mais provável. Se você está se debatendo para deixar essa crença, então vire a pergunta ao contrário e interrogue-se o seguinte: por que os números brutos de vezes em que a bolinha cai, digamos, no vermelho e no preto na roleta, *deveriam* se equilibrar à medida que renovamos os giros?

Pense no que seria necessário para fazer isso acontecer. Seria preciso que a bolinha mantivesse uma contagem de quantas vezes caiu no vermelho e no preto, detectasse qualquer discrepância e então, de algum modo, se obrigasse a cair no vermelho ou no preto para aproximar os números. Isso é pedir muito de uma simples bolinha branca ricocheteando ao acaso na roleta.

Para ser justo, superar o que os matemáticos chamam de “a falácia do jogador” significa superar a riqueza de experiências cotidianas que parecem sustentá-la. O fato é que a maioria dos nossos encontros com o acaso são mais complexos do que meros lançamentos de moedas, e facilmente podem parecer violar a lei das médias.

Por exemplo, imagine que estejamos revirando o caos que é a nossa gaveta de meias antes de sair correndo para o trabalho, à procura de um dos poucos

pares de discretas meias pretas. As chances são de que as primeiras meias sejam coloridas. Então, fazemos a coisa óbvia e as tiramos da gaveta, enquanto persistimos na busca. Agora, quem diz que a lei das médias se aplica aqui, e que uma sequência de meias coloridas não afeta as chances de se encontrar uma meia preta? Bem, isso pode parecer vagamente similar, entretanto, o que estamos fazendo é totalmente diferente de lançar uma moeda ou jogar uma bolinha na roleta. Com as meias, somos capazes de remover os resultados que não nos agradam, aumentando assim a proporção de meias pretas restantes na gaveta. Isso não é possível com eventos como um lançamento de moeda. A lei das médias não se aplica mais, porque ela diz que cada evento não afeta o seguinte.

Outra barreira que enfrentamos para aceitar a lei é que raramente lhe damos oportunidade suficiente para se revelar. Suponha que resolvamos testar a lei das médias e realizar um experimento científico apropriado envolvendo lançar uma moeda dez vezes. Poderia parecer um número razoável de tentativas; afinal, quantas vezes em geral tentamos algo antes de ficarmos convencidos de que aquilo é verdadeiro? Três vezes, talvez, meia dúzia? Na realidade, dez lançamentos não é nada perto de suficiente para demonstrar a lei das médias com alguma confiabilidade. De fato, com uma amostra tão pequena, poderíamos acabar convencidos da falácia de que os números brutos se equilibram. A matemática de cara ou coroa mostra que, em dez lançamentos, há grande chance de que a diferença entre o número de caras e o de coroas seja de 2; até há 1 chance em 4 de dar empate.

Não é de admirar que tantos de nós pensemos que “a experiência do dia a dia comprova” que os números brutos de caras e coroas se equilibram com o tempo, e não suas frequências relativas.

Conclusão

Ao tentar dar sentido a eventos aleatórios, tenha cuidado ao confiar no “senso comum” e na experiência cotidiana. Como veremos repetidamente neste livro, as leis que regem eventos aleatórios apresentam uma profusão de armadilhas para aqueles que não conhecem essas ciladas traiçoeiras.

3. O obscuro segredo do teorema áureo

OS MATEMÁTICOS ÀS VEZES alegam que simplesmente são gente como todo mundo; não são, não. Esqueça os clichês sobre bizarrices sociais e uma inclinação para roupas esquisitas; muitos matemáticos têm uma aparência perfeitamente normal. Mas todos compartilham uma característica que os distingue das pessoas comuns: uma obsessão pela prova. Não se trata de “prova” no sentido judicial nem o resultado de um experimento. Para os matemáticos, essas são coisas ridiculamente inconvincentes. Eles se referem a uma prova absoluta, garantida, *matemática*.

À primeira vista, a recusa em aceitar a palavra de alguém para alguma coisa parece bastante louvável. Mas os matemáticos insistem em aplicá-la a questões que o resto de nós consideraria obviamente verdades. Eles adoram provas rigorosas do tipo do teorema da curva de Jordan, que diz que, se você desenhar qualquer linha fechada num pedaço de papel, ela estará criando duas regiões: uma dentro da linha fechada e outra fora. Para ser justo, às vezes esse ceticismo extremo acaba se mostrando bem fundamentado. Quem adivinharia, por exemplo, o resultado da soma $1 + 2 + 3 + 4 + \text{etc. até o infinito}$? Com mais frequência, a prova confirma aquilo que os matemáticos já suspeitavam. Mas ocasionalmente uma prova de algo “óbvio” acaba se revelando impressionantemente difícil e com implicações chocantes. Dada sua reputação para mostrar surpresas, talvez não seja surpresa nenhuma que esse tipo de prova tenha surgido durante as primeiras tentativas de trazer algum rigor à teoria dos eventos aleatórios – especificamente, a definição de “probabilidade” de um evento.

O QUE SIGNIFICA “60% DE CHANCE DE CHOVER”?

Você está pensando em dar um passeio na hora do almoço, mas se lembra de ter ouvido a previsão do tempo avisar que existe uma chance de 60% de chover. Então, o que fazer? Isso depende do que você acha que significa essa chance de 60% – e há uma boa chance de não ser o que você acha. As previsões do tempo baseiam-se em modelos de computador que reproduzem a atmosfera, e, no começo dos anos 1960, os cientistas descobriram que esses modelos são “caóticos”, o que implica que até erros minúsculos nos dados que alimentam os cálculos podem produzir previsões radicalmente diferentes. Pior ainda, essa sensibilidade dos modelos muda de maneira imprevisível – tornando algumas previsões inerentemente menos confiáveis que outras. Assim, desde a década de 1990, os meteorologistas têm usado cada vez mais os chamados métodos conjuntos, fazendo dezenas de previsões, cada qual baseada em dados um pouquinho distintos, e vendo como divergem no decorrer do tempo. Quanto mais caóticas as condições, maior a divergência e menos exata a previsão final. Será que isso quer dizer que “60% de chance de chover na hora do almoço” significa que 60% da previsão conjunta mostrou chuva? Infelizmente, não: como a previsão conjunta é apenas um modelo do real, sua confiabilidade em si é incerta. Assim, o que em geral a previsão nos dá é a chamada “probabilidade de precipitação”, que leva tudo isso em conta, mais as chances de a nossa localidade realmente receber chuva. Eles alegam que essa probabilidade híbrida ajuda as pessoas a tomar melhores decisões. Talvez sim, mas em abril de 2009 o Serviço Meteorológico do Reino Unido certamente tomou uma decisão ruim ao declarar que havia “possibilidade de um verão ensolarado”. Para os versados no jargão da probabilidade, isso simplesmente significava que o modelo de computador indicara que as chances eram maiores que 50%. Contudo, para a maioria das pessoas, “possibilidade de” significa “muito provável”. Acabou que aquele foi um verão terrível, e o Serviço Meteorológico foi ridicularizado – o que é sempre uma constante certeza.

Uma das coisas mais intrigantes em relação à probabilidade é a sua natureza escorregadia, volúvel. Sua própria definição parece mudar de acordo com o que estamos pedindo dela. Às vezes parece bastante simples. Se queremos saber as chances de tirar 6 no dado, parece ok pensar nas probabilidades em termos de frequências – isto é, o número de vezes que tiramos o resultado desejado dividido pelo número total de oportunidades de que isso ocorra. Para um dado, como cada número ocupa uma das seis faces, parece razoável falar da probabilidade como a frequência a longo prazo de obter o número que queremos, que é 1 em 6. Mas o que significa falar das chances de um cavalo ganhar uma corrida? E o que os meteorologistas querem dizer quando afirmam que há 60% de chance de chover amanhã? Seguramente vai chover ou não vai? Ou será que

os meteorologistas estão tentando transmitir confiança na sua previsão? (Acontece que não é nem uma coisa nem outra – ver Box anterior.)

Os matemáticos não se sentem à vontade com esse tom vago – como mostraram quando começaram a demonstrar sério interesse no funcionamento do acaso mais ou menos 350 anos atrás. Definir o conceito de probabilidade fazia parte da sua lista de coisas a fazer. Contudo, a primeira pessoa a promover um progresso de verdade no problema viu-se recompensada com o primeiro relance do segredo obscuro sobre a probabilidade que até hoje segue de perto sua aplicação.

Nascido em Basileia, Suíça, em 1655, Jacob Bernoulli foi o mais velho da mais celebrada família matemática da história. No decorrer de três gerações, a família produziu oito matemáticos brilhantes, que, juntos, ajudaram a assentar as fundações da matemática aplicada e da física. Jacob começou a ler avidamente a então recém-emergente teoria da probabilidade na casa dos vinte anos, e ficou fascinado pelas suas potenciais aplicações em tudo, desde jogos de azar até a previsão de expectativa de vida. Mas reconheceu que havia algumas lacunas enormes na teoria, lacunas que precisavam ser preenchidas – a começar pelo significado exato de probabilidade.¹

Cerca de um século antes, um matemático italiano chamado Girolamo Cardano demonstrara a conveniência de descrever eventos regidos pelo acaso em termos da sua frequência relativa. Bernoulli decidiu fazer o que os matemáticos fazem: ver se era possível criar uma definição rigorosa. Logo percebeu, porém, que a tarefa aparentemente misteriosa gerava um imenso desafio prático. Claramente, se estamos tentando estabelecer a probabilidade de algum evento, quanto mais dados tivermos, mais confiável será nossa estimativa. Mas de quantos dados precisamos exatamente antes de dizer que “sabemos” qual é a probabilidade? Na verdade, será que esta chega a ser uma pergunta significativa de se fazer? Será que probabilidade é algo que nunca podemos saber com exatidão?

Apesar de ser um dos matemáticos mais capazes da sua época, Bernoulli levou vinte anos para responder a essas perguntas. Ele confirmou a intuição de

Cardano, de que frequências relativas são o que importa quando se quer dar sentido a eventos do acaso, como o lançamento de moedas. Ou seja, ele teve sucesso em identificar a verdadeira identidade do “tudo” em afirmações do tipo “a longo prazo tudo se equilibra”. Dessa forma, Bernoulli tinha identificado e provado a versão correta da lei das médias, que focaliza as frequências relativas, em vez de eventos individuais.

Mas isso não foi tudo. Bernoulli confirmou também o fato “óbvio” de que, quando se trata de identificar probabilidades, quanto mais dados, melhor. Especificamente, mostrou que, à medida que os dados se acumulam, o risco de as frequências medidas serem absurdamente diferentes da probabilidade real fica cada vez menor (se você acha que isso é menos convincente, parabéns: você descobriu por que os matemáticos chamam o teorema de Bernoulli de lei *fraca* dos grandes números; a versão “forte”, mais impressionante, só foi provada cerca de um século atrás).

Num sentido, o teorema de Bernoulli é a rara confirmação de uma intuição de senso comum referente a eventos regidos pelo acaso. Como ele mesmo afirmou, de maneira bastante grosseira, “mesmo a pessoa mais tola” sabe que, quanto mais dados, melhor. Mas cave um pouco mais fundo, e o teorema revela um desvio tipicamente sutil do acaso: não podemos jamais “saber” a verdadeira probabilidade com certeza absoluta. O melhor que podemos fazer é coletar tantos dados que seja possível diminuir o risco de estarmos exageradamente errados em algum nível aceitável.

Provar tudo isso foi uma façanha monumental – como o próprio Bernoulli percebeu, chamando sua prova de *theorema aureum*, “teorema áureo”. Ele estava assentando as fundações tanto da probabilidade quanto da estatística, permitindo que dados brutos sujeitos a efeitos aleatórios se transformem em percepções confiáveis.

Tendo sua predileção matemática pela prova satisfeita, Bernoulli começou a reunir seus pensamentos para sua *opus magnun*, *Ars Conjectandi*, a arte de conjecturar. Sedento de mostrar o poder prático de seu teorema, propôs-se a

aplicá-lo a problemas da vida real. Foi então que o *teorema* começou a perder um pouco de brilho.

O teorema de Bernoulli mostrava que probabilidades podem ser definidas com qualquer nível de confiabilidade – dispondo-se de dados suficientes. Assim, a pergunta óbvia era: quantos dados eram o “suficiente”? Por exemplo, se queremos saber a probabilidade de alguém com certa idade morrer no próximo ano, qual o tamanho da base de dados que precisamos para obter uma resposta que seja, digamos, 99% confiável? Para manter as coisas claras, Bernoulli usou seu teorema para atacar uma questão muito simples. Imagine um jarro enorme contendo uma mistura aleatória de pedras pretas e brancas. Suponha que nos digam que o jarro contém 2 000 pedras pretas e 3 000 brancas. A probabilidade de tirarmos uma pedra branca é, portanto, de 3 000 num total de 5 000, ou 60%. Mas, e se não conhecemos essas proporções – e portanto a probabilidade de tirar uma pedra branca? Quantas pedras precisaríamos tirar para ter confiança de estarmos bastante perto da probabilidade real?

Num típico estilo matemático, Bernoulli indicou que, antes de usarmos o teorema áureo, precisamos definir esses dois conceitos vagos de “bastante perto” e “ter confiança”. O primeiro significa exigir que os dados nos levem para dentro de, digamos, mais ou menos 5% da probabilidade real, ou mais ou menos 1%, ou ainda mais perto. Confiança, por outro lado, concentra-se na frequência com que atingimos esse nível de precisão. Podemos resolver que queremos ter confiança de atingir esse padrão 9 vezes em 10 (“90% de confiança”) ou 99 vezes em 100 (“99% de confiança”), ou uma confiança ainda maior.² O ideal, obviamente, é ter 100% de confiança, mas, como deixa claro o teorema áureo, em fenômenos afetados pelo acaso essa certeza divina não é atingível.

O teorema áureo parecia captar a relação entre precisão e confiança para o problema das pedras coloridas tiradas ao acaso não só de um jarro, mas de *qualquer* jarro. Então Bernoulli pediu-lhe que revelasse o número de pedras que deveriam ser retiradas do jarro para haver 99,9% de confiança de ter identificado as proporções relativas de pedras brancas e pretas ali contidas, com uma margem de mais ou menos 2%. Inserindo esses números em seu teorema, ele girou a

manivela matemática... e surgiu uma resposta chocante. Se o problema precisasse ser resolvido tirando pedras ao acaso, seria necessário examinar mais de 25 500 pedras antes que as proporções relativas das duas cores pudessem ser definidas pelas especificações de Bernoulli.

Esse não era apenas um número tristemente grande, era grande num nível ridículo. Sugeria que a amostragem aleatória era um meio irremediavelmente ineficiente de avaliar proporções relativas, pois, mesmo num jarro com apenas alguns milhares de pedras, seria necessário repetir o processo de examinar as pedras mais de 25 mil vezes para obter a verdadeira porcentagem segundo o padrão de Bernoulli. Estava claro que seria muito mais rápido tirar as pedras e contá-las. Historiadores ainda discutem sobre o que Bernoulli teria pensado de sua estimativa;³ parece que o consenso foi “decepção”. O certo é que, depois de anotar a resposta, ele adicionou mais algumas linhas ao seu trabalho, e então parou. *Ars Conjectandi* definiu sem ser publicado até 1713, oito anos após a morte de seu autor. É difícil evitar a suspeita de que Bernoulli perdera a confiança no valor prático do teorema áureo. Sabe-se que ele estava ansioso para aplicá-lo a problemas muito mais interessantes, inclusive para resolver disputas legais em que se necessitava uma evidência para deixar o caso “para além da dúvida razoável”. Bernoulli parece ter manifestado decepção nas implicações de seu teorema numa carta ao distinto matemático alemão Gottfried Leibniz, na qual admitia não conseguir achar “exemplos adequados” dessas aplicações para o teorema.

Seja qual for a verdade, sabemos agora que, embora o teorema de Bernoulli tivesse lhe fornecido a compreensão conceitual que ele buscava, ainda era necessária alguma carga matemática turbinada antes de ele ser usado em problemas da vida real. Essa carga foi aplicada após a morte de Bernoulli pelo brilhante matemático francês (e amigo de Isaac Newton) Abraham de Moivre – permitindo que o teorema funcionasse com número bem menor de dados.⁴ Todavia, a fonte real do problema não residia tanto no teorema quanto nas expectativas que Bernoulli alimentava em relação a ele. Os níveis de confiança e precisão que ele impunha lhe pareciam razoáveis, mas eram rigorosos demais.

Mesmo usando a versão moderna de seu teorema, estabelecer a probabilidade para os padrões que Bernoulli determinou exige cerca de 7 000 pedras aleatoriamente tiradas do jarro e com a cor anotada – o que ainda é uma quantidade enorme.

É estranho que Bernoulli não tivesse feito a coisa óbvia e retrabalhado seus cálculos com exigências bem menores quanto à precisão e à confiança. Pois mesmo na sua forma original, o teorema áureo mostra que isso tem um impacto significativo na quantidade de dados requeridos; usando a versão moderna, o impacto é bastante drástico. Tomando-se o nível de confiança de 99,9% estabelecido por Bernoulli, mas flexibilizando-se o nível de precisão de mais ou menos 2% para 3%, corta-se o número de observações para menos da metade, algo em torno de 3 000. Outra alternativa é manter o nível de precisão em 2% mas reduzir o nível de confiança para 95%, o que corta o número de observações ainda mais, para algo em torno de 2 500 – apenas 10% da quantidade estimada por Bernoulli. Fazendo-se as duas coisas – um pouco menos de precisão, um pouco menos de confiança –, o número despenca de novo, para algo em torno de mil.

Esse é um valor bem menos exigente que o número alcançado por Bernoulli, embora, reconhecidamente, tenhamos de pagar um preço em termos de confiabilidade do nosso conhecimento. Talvez Bernoulli tivesse resistido à ideia de baixar tanto seus padrões; infelizmente, nunca saberemos.

Hoje, 95% tornou-se o padrão de fato para os níveis de confiança numa profusão de disciplinas orientadas por dados, da economia à medicina. Organizações de pesquisa combinaram essa confiança com uma precisão de mais ou menos 3% para chegar ao tamanho-padrão da amostra de pesquisa, de aproximadamente mil. Todavia, embora possam ser bastante usados, nunca devemos esquecer que esses padrões baseiam-se no pragmatismo, e não em algum consenso grandioso do que constitui “uma prova científica”.

Conclusão

O segredo obscuro que está à espreita no teorema áureo de Bernoulli é que, quando se tenta avaliar os efeitos do acaso, uma certeza do tipo divina é inatingível. Em vez disso, geralmente deparamos com um meio-termo entre juntar mais evidência ou reduzir nosso padrão de conhecimento.

4. A primeira lei da ausência de leis

O VERDADEIRO SIGNIFICADO da lei das médias tem sido deturpado e mal compreendido de uma forma tão grave e com tamanha frequência que os especialistas em probabilidade tendem a evitar o termo. Eles indiscutivelmente preferem expressões ainda menos úteis, como lei fraca dos grandes números – que soa como regra pouco confiável acerca de multidões. Então, em vez disso, vamos dividir a lei das médias nas concepções que a compõem e chamá-las de “leis da ausência de leis”. A primeira delas concentra-se na melhor forma de pensar a respeito de eventos que envolvam um elemento de acaso.

A PRIMEIRA LEI DA AUSÊNCIA DE LEIS

Ao tentar dar sentido a eventos envolvendo o acaso, ignore os números brutos. Em vez disso, focalize a atenção na frequência relativa – isto é, a frequência com que eles ocorrem dividida pela frequência com que teriam oportunidade de ocorrer.

A primeira lei da ausência de leis nos adverte para termos cautela diante de afirmações que se baseiam exclusivamente em números brutos de eventos. Isso a torna especialmente proveitosa quando confrontada, por exemplo, com a cobertura de mídia sobre pessoas que apresentam efeitos colaterais a algum novo tratamento, ou com os prêmios da loteria numa cidade específica. Essas histórias são caracteristicamente acompanhadas por fotos de vítimas trágicas ou felizardos ganhadores. Não há dúvida do poder dessas matérias. Até um só caso chocante na vida real pode deflagrar mudanças históricas na elaboração de políticas – como sabe muito bem qualquer pessoa que tenha passado pela segurança do aeroporto nos Estados Unidos depois do 11 de Setembro. E às vezes a resposta

apropriada é essa mesma. Mas basear uma decisão num punhado de casos geralmente é uma ideia muito ruim.

O perigo é que os casos parecem típicos, quando de fato não são nada disso. Realmente, às vezes eles são tão chocantes porque estão “fora da curva” – são produto de confluências do acaso extremamente raras.

A primeira lei da ausência de leis mostra que podemos evitar essas ciladas concentrando-nos nas *frequências relativas*: o número bruto de eventos dividido pelo número relevante de oportunidades para que eles ocorram.

Vamos aplicar a lei a um exemplo da vida real: a decisão tomada em 2008, pelo governo do Reino Unido, de vacinar meninas pré-adolescentes contra o HPV, o vírus responsável pelo câncer de colo do útero. Saudou-se esse programa nacional pela potencialidade de salvar a vida de centenas de mulheres por ano. No entanto, pouco depois de lançado, a mídia parecia ter uma evidência inquestionável de que aquela era uma visão perigosamente otimista. Foi relatado o trágico caso de Natalie Morton, menina de catorze anos que morreu poucas horas depois de ter recebido a vacina. As autoridades de saúde responderam conferindo os estoques e retirando o lote suspeito. Entretanto, isso não bastou: queriam que se abandonasse a vacinação em massa. Isso era algo razoável? Alguns insistiam, invocando o chamado princípio da precaução, que, na sua forma menos sofisticada, redundava em “Melhor prevenir que remediar”. O perigo aqui está em resolver um problema criando outro. Interromper o programa eliminaria qualquer risco de morte entre as participantes, mas ainda resta o problema de como encarar o câncer de colo do útero.

Depois há o risco de cair numa cilada que merece ser mais bem conhecida (e que encontraremos novamente neste livro). Os lógicos a chamam de falácia *post hoc, ergo propter hoc* – expressão latina que quer dizer “depois disso, portanto, por causa disso”. No caso da morte de Natalie, a cilada está em assumir que, por ela ter morrido *depois* de ser vacinada, a vacina deve ter sido a causa. Sem dúvida alguma, causas verdadeiras sempre precedem seus efeitos, mas inverter a lógica representa um perigo: as pessoas em acidentes de carro costumam pôr o

cinto de segurança antes de iniciar a viagem, mas isso não significa que pôr o cinto cause o acidente.

Mas vamos admitir o pior: que a morte de Natalie realmente tenha sido causada por uma reação adversa à vacina. A primeira lei da ausência de leis nos diz que a melhor maneira de dar sentido a esses eventos é focalizar não os casos individuais, e sim as proporções relevantes. Na época da morte de Natalie, 1,3 milhão de garotas haviam recebido a mesma vacina. Isso quer dizer que a frequência relativa desse tipo de evento era em torno de 1 em 1 milhão. Foi o que persuadiu o governo do Reino Unido, diante dos protestos dos manifestantes antivacinação, a retomar o programa uma vez retirado o lote suspeito. Essa era a resposta racional no caso de Natalie ter sido de fato vítima de uma reação rara à vacina.

Acontece que não foi isso o que aconteceu: a mídia realmente caíra na armadilha do *post hoc, ergo propter hoc*. No inquérito sobre a morte da menina, veio à tona que Natalie tinha um tumor maligno no tórax, e sua morte não teve nenhuma ligação com a vacina. Mesmo assim, a primeira lei mostra que as autoridades haviam adotado a abordagem correta retirando apenas o lote suspeito, em vez de abandonar todo o programa.

Claro que a primeira lei não é uma garantia que leve diretamente à verdade. Natalie poderia ter sido o caso zero de uma reação à vacina nunca detectada durante os testes. Evidentemente, era certo examinar as causas do caso em busca de evidências de que aquilo poderia ocorrer de novo. O papel da primeira lei está em nos impedir de ficar exageradamente impressionados com os casos individuais e, em vez disso, focalizar nossa atenção nas frequências relativas, colocando dessa forma esses casos em seu contexto correto.

Aqui há mais lições genéricas para gerentes, administradores e políticos determinados a fazer “melhorias” após um punhado de eventos únicos. Se ignorarem a primeira lei da ausência de leis, eles se arriscam a tomar atitudes para lidar com eventos excessivamente raros. Pior ainda, baseando a “melhoria” numa quantidade pequena de casos, eles podem decidir testá-la num conjunto de dados igualmente diminuto, mais uma vez se concentrando nos números brutos,

e não nas frequências relativas, e chegando assim a conclusões absolutamente erradas. Pode ser qualquer tema, desde uma inundação de queixas de clientes até uma sugestão da equipe sobre, digamos, um jeito novo de fazer as coisas. Tudo isso tende a começar com alguns casos isolados que podem ou não ser significativos. Mas o primeiro passo para descobrir é colocá-los no contexto adequado – transformando-os em suas apropriadas frequências relativas.

Às vezes dar sentido aos eventos requer uma *comparação* de frequências relativas. No fim dos anos 1980, a empresa privada de defesa GEC-Marconi, com sede no Reino Unido, tornou-se o centro da cobertura da mídia após uma leva de mais de vinte suicídios, mortes e desaparecimentos em sua equipe técnica. Começaram a surgir teorias conspiratórias, alimentadas pelo fato de que algumas das vítimas trabalhavam em projetos sigilosos. Ainda que estes gerem histórias intrigantes, a primeira lei nos diz para ignorar os casos isolados e, em vez disso, focar as frequências relativas – nesse caso, uma comparação entre a frequência relativa de eventos estranhos na Marconi e os casos que seriam de esperar na população geral. Isso imediatamente concentra a atenção no fato de que a GEC-Marconi era uma empresa enorme, empregando mais de 30 mil funcionários, e que as mortes haviam se espalhado por um período de oito anos. As mortes e os desaparecimentos “misteriosos” não eram tão surpreendentes, dado o tamanho da empresa. Foi a essa conclusão que chegou a posterior investigação policial, embora teorias conspiratórias persistam até hoje.

Para ser justo, a importância de comparar frequências relativas está começando a crescer na mídia. Em 2010, a France Telecom invadiu as manchetes com um número de suicídios do tipo da GEC-Marconi: trinta, entre 2008 e 2009. A história voltou a ganhar destaque em 2014, quando a empresa – agora chamada Orange Telecom – assistiu ao ressurgimento de suicídios, com dez em apenas poucos meses. Dessa vez, a explicação *du jour* foi o estresse relacionado ao trabalho. Mas, em contraste com as reportagens dos casos da GEC-Marconi, alguns jornalistas propuseram a questão-chave induzida pela primeira lei: será que a taxa de suicídios, e não apenas os números brutos, é

realmente tão anormal – uma vez que se trata de uma empresa enorme, com cerca de 100 mil funcionários?

O ESTRANHO CASO DO TRIÂNGULO DAS BERMUDAS

A primeira lei é especialmente útil quando se tenta dar sentido a explicações sinistras e a teorias conspiratórias. Peguemos o caso bem conhecido de desaparecimento de navios e aviões sobre uma região do Atlântico ocidental conhecida como Triângulo das Bermudas. Da década de 1950 em diante, houve incontáveis relatos de que coisas ruins acontecem com aqueles que entram nessa área em forma de triângulo entre Miami, Porto Rico e a ilha de mesmo nome. Muitas teorias têm se apresentado para explicar os eventos, desde ataques de óvnis até ondas maléficas. Mas a primeira lei da ausência de leis nos diz para não nos concentrarmos nos números brutos de desaparecimentos “misteriosos” (que podem ou não ter ocorrido), e comparar sua frequência relativa com o que seria de esperar em qualquer parte correspondente do oceano. Faça isso, e surge algo de arrepiar: é inteiramente possível que todos os desaparecimentos não explicados tenham realmente ocorrido. Isso porque dezenas de milhares de navios e aviões passam todo ano por essa vasta área, de cerca de 1 milhão de quilômetros quadrados de mar e espaço aéreo. Mesmo que se incluam todos os relatos estranhos de casos não explicados, descobre-se que o Triângulo das Bermudas não está sequer entre as dez principais zonas de perigo oceânico. Decerto os empertigados atuários da mundialmente famosa seguradora Lloyd's de Londres não se perturbam com os números brutos de eventos supostamente “misteriosos” na região. Eles não cobram prêmios de seguro mais caros pelo risco de se aventurar nessa área.

No entanto, isso suscita a questão traiçoeira que muitas vezes emerge quando se tenta aplicar a primeira lei: qual a frequência relativa apropriada para se usar na comparação? No caso da Orange Telecom, será a taxa nacional de suicídios (sabidamente alta na França, mais ou menos 40% acima da média da União Europeia), ou algo mais específico, como a taxa entre faixas etárias particulares (suicídio é a principal causa de morte entre pessoas de 25-34 anos na França) ou talvez grupos socioeconômicos? Ainda não há uma conclusão sobre o caso da Orange Telecom; embora isso possa ser uma simples anomalia estatística passageira, há quem insista em que a verdadeira explicação é o estresse no local de trabalho. É muito possível que nunca se saiba a verdade.

Qualquer que seja a realidade, a primeira lei nos diz onde começar para dar sentido a essas questões. E também faz uma predição: qualquer coisa que abranja gente suficiente – desde uma campanha governamental de saúde até empregos numa multinacional – tem a capacidade de gerar histórias que dão manchetes, respaldadas por casos isolados da vida real, que significam menos do que parecem.

Tente você mesmo. Da próxima vez que ouvir falar de alguma campanha nacional que seja boa, em geral, mas que possa ter efeitos colaterais perniciosos para algumas pessoas – por exemplo, uma campanha de medicação em massa –, tome nota, espere pelas histórias de horror e ponha em funcionamento a primeira lei.

Conclusão

Eventos regidos pelo acaso podem nos chocar pela aparente improbabilidade. A primeira lei da ausência de leis nos diz para olhar além dos números brutos desses eventos e focalizar suas frequências relativas – o que nos dá a possibilidade de lidar com o evento. Se eventos de baixa probabilidade podem ocorrer, eles ocorrerão – quando tiverem oportunidade suficiente.

5. Quais são as chances *disso*?

SUE HAMILTON ESTAVA trabalhando com uma papelada no seu escritório em Dover, em julho de 1992, quando deparou com um problema. Achou que seu colega, Jason, talvez soubesse como resolvê-lo, mas, como ele tinha ido para casa, resolveu lhe telefonar. Descobriu o número do telefone no quadro de avisos do escritório. Depois de se desculpar por incomodá-lo em casa, começou a explicar o problema, porém, mal tinha começado, Jason a interrompeu para avisar que não estava em casa. Estava numa cabine pública de telefone. O aparelho começara a tocar justo quando ele vinha passando; Jason parou e resolvera atender. Espantosamente, aquele número no quadro de avisos não era em absoluto o de Jason. Era o número do seu registro de empregado – que por acaso era idêntico ao número do telefone da cabine pela qual ele estava passando no momento em que Sue ligara.

Todo mundo adora histórias de coincidências. Elas parecem insinuar conexões invisíveis entre eventos e nós, governadas por leis misteriosas. E é verdade. Há uma miríade de conexões invisíveis entre nós, mas elas são invisíveis basicamente porque não saímos por aí procurando. As leis que as governam também são misteriosas – porém, mais uma vez, é essencialmente porque poucas vezes alguém nos fala sobre elas.

Coincidências são manifestações da primeira lei da ausência de leis, mas com uma pequena diferença. A lei nos conta o que fazer para dar sentido a eventos regidos pelo acaso, enquanto as coincidências nos advertem sobre quanto pode ser difícil fazer isso.

Quando confrontada com uma coincidência “espantosa”, a primeira lei nos diz para começar nos perguntando sobre sua frequência relativa – ou seja, o número de vezes que essa coincidência espantosa poderia ocorrer dividido pelo número de oportunidades que os eventos têm de ocorrer. Para uma coincidência

realmente espantosa, é de esperar que a estimativa da probabilidade do evento fosse impressionantemente baixa. Mas quando tentamos aplicar a lei a coincidências como o telefonema de Sue Hamilton, acabamos em apuros.

Como começamos a estimar o número desses eventos espantosos, ou o número de oportunidades em que eles podem se dar? Para começar, o que quer dizer “espantoso”? Decididamente, não é algo que possamos definir de modo objetivo, o que por sua vez representa que estamos em solo movediço ao insistir que vivenciamos algo significativo em si mesmo. O grande e saudoso físico ganhador do Prêmio Nobel Richard Feynman ressaltou esse traço comum das coincidências com um exemplo tipicamente pé no chão. Durante uma palestra sobre como dar sentido à evidência, disse à plateia o seguinte:

Sabem, esta noite me aconteceu uma coisa muito impressionante. Eu estava vindo para cá, a caminho da palestra, e entrei pelo estacionamento. Vocês não imaginam o que aconteceu. Vi um carro com a placa ARW 357. Podem imaginar? De todos os milhões de placas de carro neste estado, qual a chance de eu ver essa placa específica esta noite? Impressionante!

Então, há o fato incômodo de que em geral decidimos que uma coincidência é “espantosa” só depois que a vivenciamos, tornando nossa avaliação acerca de seu significado *post hoc*, e potencialmente enganosa. Há um esquete do Monty Python baseado na lenda de Guilherme Tell que capta perfeitamente os perigos de uma racionalização *post hoc*. O quadro mostra uma multidão de pessoas reunidas em torno do nosso mencionado herói, enquanto ele faz cuidadosa pontaria na maçã colocada sobre a cabeça de seu filho – e acerta! A multidão ovaciona devidamente... e nós também nos sentimos impressionados, até que a câmera vai recuando para revelar o filho de Tell crivado de setas, de todas as tentativas anteriores fracassadas de acertar. A habilidade de Tell só parece espantosa se ignorarmos todos os fracassos; é isso que acontece com as coincidências. Na realidade, elas ocorrem o tempo todo à nossa volta, mas a esmagadora maioria é tediosa e insignificante. De vez em quando localizamos algo que decidimos ser equivalente a uma seta partindo a maçã ao meio – e declaramos que é surpreendente, espantoso ou até misterioso, ignorando cuidadosamente a miríade de eventos menos interessantes.

Tudo isso fala do fato de que nós, seres humanos, somos inatos buscadores de padrões, propensos a ver sentido em ruídos sem significado nenhum. Sem dúvida nossos ancestrais habitantes das cavernas se beneficiavam errando pelo lado de excesso de cautela, e se escondiam se algo se parecesse vagamente com um predador. Mas isso pode escorregar facilmente para aquilo que os psicólogos chamam de apofenia: a predileção por enxergar padrões onde eles não existem. Todos nós estamos especialmente propensos a uma forma específica de apofenia conhecida como pareidolia. Vez ou outra a mídia reporta argumentos sobre formações de nuvens “miraculosas”, marcas chamuscadas em torradas ou traços em mapas do Google que supostamente se parecem com Cristo, madre Teresa ou Kim Kardashian. É difícil discordar de que isso de fato aconteça. O que concluimos sobre esses “milagres” depende, se julgamos que as chances de eles ocorrerem por mera casualidade são incrivelmente pequenas. Se aplicarmos a primeira lei da ausência de leis, temos de confrontar o fato de que o cérebro tem uma miríade de maneiras de criar um rosto a partir de uma espiral aleatória.

Um dos casos mais conhecidos de pareidolia gira em torno do chamado Rosto de Marte. Em 1976, uma das sondas da Nasa no “planeta vermelho” enviou uma foto que parecia exibir a imagem de um alienígena no planeta. A figura provocou controvérsias durante 25 anos, com a maioria dos cientistas desconsiderando-a, como uma grande bobagem. Alguns tentaram estimar as chances de obter um rosto tão realista por puro acaso, mas acabaram atolados em discussões sobre os números que haviam introduzido em seus cálculos das frequências relativas. Finalmente, em 2001, a verdade foi revelada por imagens bem-definidas tiradas pela sonda Mars Global Surveyor. As imagens mostravam que o “rosto” era na verdade uma formação rochosa, exatamente como argumentavam os céticos.

Ao tentar dar sentido a uma coincidência, é fácil subestimar como é comum o evento “espantoso” – no mínimo por definir quão espantoso ele é só *depois* de vê-lo, ou, na realidade, de trapacear.

COMO PREDIZER COINCIDÊNCIAS

Uma das demonstrações mais estarrecedoras das leis da probabilidade é o chamado paradoxo do aniversário: são suficientes apenas 23 pessoas para haver uma chance maior que 50:50 de que duas delas façam aniversário no mesmo dia. No entanto, você não precisa de um grupo tão grande para demonstrar essas coincidências: uma reunião aleatória de cinco pessoas dá uma chance bem razoável de que pelo menos duas tenham o mesmo signo astrológico (ou tenham nascido no mesmo mês, se você não for um virginiano racional e preferir exemplo menos bobo). A razão de se precisar de tão pouca gente é que você está pedindo *qualquer* igualdade de data entre todos os diferentes modos de formar pares com duas pessoas quaisquer do grupo – o que resulta num número surpreendentemente grande: podem se formar 253 pares com 23 pessoas. Essa falta de especificidade é a chave: se você quiser uma coincidência exata com o *seu* aniversário, vai precisar de uma multidão de mais de 250 pessoas para obter chance maior que 50:50. Sendo menos exigente e procurando dois aniversários *quaisquer* com diferença de um dia a mais ou a menos, as chances aumentam tremendamente: de fato, há 90% de chance de encontrar essa “quase” coincidência entre os jogadores de qualquer partida de futebol.¹

Conclusão

As coincidências nos surpreendem porque pensamos que elas são muito improváveis, logo, não podem acontecer “por mera casualidade”. A primeira lei da ausência de leis nos adverte dos perigos de subestimar as chances de coincidência resolvendo nós mesmos o que contamos como “espantoso”.

6. Pensar de modo independente não inclui gema de ovo

EM SETEMBRO DE 2013, John Winfield estava na cozinha de sua casa em Breadsall, Derbyshire, quando percebeu que precisava de alguns ovos. Deu um pulo até a mercearia, voltou com seis ovos e começou a quebrá-los. Para sua surpresa, o primeiro tinha uma gema dupla – algo que ele nunca tinha visto antes na vida. Então quebrou outro, e viu outra gema dupla. Perplexo, continuou quebrando os ovos, e descobriu que todos tinham gemas duplas, inclusive o último – que deixou cair no chão, de tão agitado.

O espantoso caso das seis gemas duplas chegou ao conhecimento de jornalistas, que prestativamente fizeram os cálculos para mostrar quanto era improvável o evento. Segundo o Serviço Britânico de Informação sobre Ovos, em média, apenas 1 entre 1 000 ovos produzidos tem gema dupla. E isso incentivou os repórteres a pegar suas calculadoras e mais algumas noções vagas sobre como lidar com as probabilidades. Eles estimaram que, se havia 1 chance em 1 000 de obter uma gema dupla, a chance de obter 6 devia ser 1 em 1 000 multiplicada por si mesma 6 vezes, ou 1 em 1 000 000 000 000 000 (1 em 1 quintilhão, ou 1 em 1 bilhão de bilhões). Trata-se de um número astronômico: implica que, para presenciar apenas uma vez o que o sr. Winfield viu, seria preciso ter aberto uma caixa de ovos por segundo desde o nascimento do Universo.

Entretanto, alguns jornalistas perceberam que havia algo não confiável nesse raciocínio. Para começar, o sr. Winfield nem de longe era o primeiro desde o big bang a relatar tal evento. Uma rápida consulta na internet revelou diversos relatos similares, inclusive um caso idêntico de seis gemas duplas encontradas na Cúmbria três anos antes. O colunista de ciência Michael Hanlon, do *Daily Mail*, levantou dúvidas sobre a proporção 1 em 1 000 usada nos cálculos.¹ Assinalou

que as chances de obter gemas múltiplas dependiam fortemente da idade das galinhas: as galinhas jovens têm uma probabilidade 10 vezes maior de produzi-las. Assim, ainda que o número 1 em 1 000 fosse verdadeiro em média, a proporção de gemas duplas para granjas com aves mais jovens podia ser facilmente de 1 em 100 – aumentando em pelo menos 1 milhão as chances de obter uma leva de 6 nessas granjas.

Essa, porém, não pode ser toda a explicação, pois ainda deixa as chances de obter gemas duplas em algo por volta de 1 em 1 bilhão. Todo ano é consumido no Reino Unido o equivalente a cerca de 2 bilhões de caixas de meia dúzia; logo, mesmo com as chances imensamente ampliadas, ainda seria esperável ouvir cerca de dois casos por milênio, não dois em mais ou menos três anos. Quando um cálculo dá uma resposta loucamente incorreta como essa, isso é sinal de que há alguma coisa fundamentalmente errada em suas premissas. E a grande premissa feita aqui é de que as probabilidades de cada evento ocorrer separadamente podem ser multiplicadas entre si. As leis da probabilidade mostram que isso só é permitido se os eventos em questão – nesse caso, a descoberta de gemas duplas – forem independentes um do outro, de modo que não tenhamos de fazer nenhuma correção relativa a alguma influência externa.

A noção de que os eventos são independentes corre nas profundezas da teoria das probabilidades. Muitas manifestações de acaso em “livros-texto” – lançamentos repetidos de uma moeda, digamos, ou o rolar de dados – são de fato independentes; não há motivo para desconfiar que um dos eventos deva influenciar algum outro. Contudo, quando a premissa de independência mantiver a matemática simples, nunca devemos perder de vista o fato de que ela não passa exatamente disto: uma premissa. Às vezes é uma premissa que podemos construir com segurança – quando tentamos dar sentido à lendária “maré de azar” do jogador de críquete Nasser Hussain, em 2001, quando ele perdeu a disputa de cara ou coroa catorze vezes seguidas. Ainda que as chances de isso ocorrer sejam de cerca de 1 em 16 000, não há necessidade de desconfiar de nada estranho; quando se pensa em quantos excelentes jogadores de críquete lançaram moedas nas últimas décadas, esse é um evento que fatalmente iria

acontecer um dia. Mas com demasiada frequência a premissa de independência não é sequer remotamente justificável. Vivemos num mundo bagunçado, interligado, atravessado por conexões, ligações e relações. Algumas resultam das leis da física, algumas da biologia, algumas da psicologia humana. Qualquer que seja a causa das conexões, assumir alegremente que elas não existem pode nos meter em apuros. De fato, as consequências são sérias a ponto de merecer outra lei da ausência de leis.

A SEGUNDA LEI DA AUSÊNCIA DE LEIS

Ao tentar compreender sequências de eventos aparentemente “aleatórios”, não assuma de modo automático que eles são independentes. Muitos eventos no mundo real não o são – e assumir que sejam pode levar a estimativas muito enganosas acerca das chances de observar essas “sequências”.

Aplicar a segunda lei à história das gemas duplas significa pensar nas maneiras pelas quais o fato de encontrar um ovo desses numa caixa pode estar ligado a encontrar outros na mesma caixa. Como vimos, uma dessas maneiras é que o conteúdo da caixa possa ter vindo de galinhas jovens, propensas a produzir gemas duplas. Depois, a possibilidade de que os ovos de gema dupla sejam agrupados pelos embaladores de ovos, aumentando a chance de obter uma caixa cheia deles. Mais uma vez, sabe-se que isso ocorre: ovos de gema dupla tendem a ser relativamente grandes e a se destacar entre os ovos pequenos produzidos por galinhas jovens – assim, tendem a ser embalados juntos. Alguns supermercados chegam a fazer questão de que os ovos com a possibilidade de ter gema dupla estejam na mesma caixa.

Há, portanto, bases sólidas para se pensar que achar um ovo de gema dupla aumenta as chances de se encontrar outro na mesma caixa – e, portanto, para rejeitar a ideia de independência e a colossal improbabilidade aí implícita. Como a primeira lei, a segunda lei tem uma miríade de usos – inclusive dar sentido a coincidências aparentemente misteriosas. Tomemos o relato bizarro de como o

desastre do *Titanic*, em abril de 1912, foi previsto em detalhes assustadoramente acurados por um livro escrito catorze anos antes. No conto “Futilidade”, publicado em 1898, o escritor americano Morgan Robertson conta a história de John Rowland, marinheiro a bordo do maior navio já construído, que afunda com uma enorme perda de vidas após se chocar contra um iceberg no Atlântico Norte numa noite de abril. E o nome do navio? *SS Titan*. Os paralelos tampouco param aí. A embarcação de Robertson tinha cerca de 240 metros de comprimento, mais ou menos o mesmo tamanho do *Titanic*, e era descrita como “não afundável”, carregando menos da metade dos botes salva-vidas necessários para os viajantes a bordo. E até foi atingido do mesmo lado: estibordo.

Sem dúvida essa é uma lista impressionante de coincidências, e poderia levar alguém a se perguntar se Robertson baseara seu livro numa premonição. Talvez sim, mas a boa aposta está no fato de que seu enredo é uma demonstração de como as coincidências emergem se os eventos não são independentes. Quando “Futilidade” foi publicado, já estava em andamento uma corrida para construir navios de passageiros colossais, provocada pela competição internacional para ganhar a Blue Riband, a Flâmula Azul, prêmio concedido ao mais rápido transatlântico de passageiros. Na década final do século XIX, os maiores navios mediam de 170 metros a bem mais de 200 metros de comprimento – e os 240 metros não estavam fora de cogitação. Quanto ao que podia causar estragos nesses leviatãs, os icebergs já eram uma reconhecida ameaça. Como o era também a inadequada provisão de botes salva-vidas: já houvera advertências de que os regulamentos tinham fracassado na tarefa de se manter em compasso com o rápido aumento no tamanho dos navios. Claro que a adivinhação correta do lado atingido pelo iceberg foi um simples chute de 50:50 de chance. Menos surpreendente é a escolha de Robertson do nome de seu malfadado navio. Em busca de algo evocativo para uma embarcação colossal, *SS Titan* obviamente tem mais probabilidade de aparecer numa lista de candidatos que, digamos, *SS Midget*.^a

Em suma, o objetivo de Robertson em redigir um conto trágico porém plausível sobre um leviatã malfadado mais ou menos o compeliu a incluir

eventos e características não muito distantes das do *Titanic*. Uma escolha aleatória simplesmente não teria feito sentido narrativo.

Conclusão

Manifestações de casualidade que aparecem em livros-texto, como lançamentos de moedas, podem ser consideradas independentes. Mas, no mundo real, muitas vezes essa é uma premissa perigosa, mesmo com sequências de eventos aparentemente raros. A segunda lei da ausência de leis nos adverte contra assumir essa independência de modo automático ao estimar as chances de tal conjunto de coincidências.

^a *Midget*: em inglês, anão, gnomo. (N.T.)

7. Lições aleatórias da loteria

DESDE QUE COMEÇOU, em 1988, a loteria estadual da Flórida já entregou mais de US\$ 37 bilhões em prêmios, criou mais de 1 300 milionários e pagou a universidade de 650 mil estudantes. Contudo, em 21 de março de 2011, transformou uma porção de moradores do estado em adeptos das teorias da conspiração. Após anos de suspeitas, naquela noite eles acreditaram ter finalmente obtido a prova da razão de nunca terem recebido nada apesar dos anos de tentativa: a loteria era uma armação. Toda noite, sete dias por semana, a loteria faz o sorteio Fantasy 5, em que 36 bolas são colocadas numa máquina randomizadora e cinco bolas vencedoras são escolhidas ao acaso. Ou pelo menos é isso que alegam os organizadores. Mas, naquele dia de 2011, ficou óbvio que havia uma armação. À medida que as bolas saltavam da máquina, tornou-se evidente que o processo era tudo, menos aleatório: os números ganhadores foram 14, 15, 16, 17, 18. Os apostadores da pesada na loteria sabiam que a probabilidade de ganhar o grande prêmio com qualquer seleção aleatória de números era em torno de 1 em 377 000, então estava claro que algo muito suspeito tinha acontecido.

Na realidade, ocorrera uma coisa extremamente comum: uma demonstração de que a maioria de nós tem uma compreensão menos que perfeita do que é realmente a aleatoriedade.

Todos nós gostamos de pensar que é possível aprender com a experiência. E, considerando como os eventos aleatórios são comuns no nosso mundo, você vai pensar que as pessoas perceberiam com muita facilidade o que a aleatoriedade coloca no nosso caminho. Não poderia estar mais errado. Solicitadas simplesmente a definir aleatoriedade, as pessoas tipicamente mencionam características como “não ter causa ou motivo” e “ausência de padrões” – o que não é tão ruim, pelo menos até certo ponto. Mas quando são solicitadas a aplicar

essas percepções intuitivas a problemas da vida real, as coisas começam a desandar.

Na década de 1970, o psicólogo Norman Ginsburg, da Universidade McMaster, no Canadá, realizou estudos para ver quanto as pessoas são boas em executar a tarefa aparentemente simples de escrever listas de 100 dígitos aleatórios. A maioria dos participantes apareceu com sequências bem embaralhadas de dígitos, poucos deles repetidos, ou sequências de números consecutivos, ou qualquer outro padrão numérico. Em outras palavras, fizeram o melhor possível para garantir que todo dígito tivesse sua “cota justa” de presença numa sequência que, de outra maneira, estaria destituída de padrões. No processo, inadvertidamente demonstraram uma concepção errônea fundamental sobre a aleatoriedade.

É verdade que não há causa ou motivo para a aleatoriedade: por definição, ela não pode ser resultado de qualquer processo previsível. E também é verdade que ela não tem padrões. O problema é que isso só é algo garantido em escalas gigantescas (de fato, estritamente falando, infinitas). Em qualquer outra escala, a falta de causa ou motivo de aleatoriedade é inteiramente capaz de conter sequências padronizadas longas o bastante para *parecer* significativas. Todavia, quando solicitados a criarmos nós mesmos alguma aleatoriedade, não podemos resistir a tentar reproduzir a natureza sem padrões da aleatoriedade infinita, mesmo nas manifestações mais breves da coisa.

Fica claro que aquilo de que precisamos é uma exposição regular a intervalos breves de aleatoriedade, de modo a termos uma sensação de como ela é em tais escalas. Felizmente, isso se consegue com facilidade – de fato, milhões de pessoas o fazem inconscientemente no mundo todo várias vezes por semana. Chama-se assistir aos sorteios da loteria na TV.

Muitos países têm loterias nacionais como meio de arrecadar dinheiro para boas causas. A maioria das pessoas assiste aos sorteios simplesmente para ver se ganhou algum prêmio – o que, considerando-se que a chance é tipicamente 1 em milhões, em geral é um exercício de futilidade. Contudo, há algo a se dizer mesmo para aqueles que não compraram nenhum bilhete de loteria, mas

sintonizam vez por outra o canal do sorteio para ver o que a aleatoriedade pode fazer – e observar os números produzirem algo que, de modo suspeito, parece um padrão.

Muitas loterias (inclusive, até recentemente, a loteria nacional do Reino Unido) são do tipo “6 em 49”; ou seja, ganhar significa adivinhar corretamente as seis bolas sorteadas entre as 49 colocadas numa máquina randomizadora. Isso não parece muito difícil; é estranhamente tentador estimar que a chance de acertar o conjunto correto de seis bolas é de 6 em 49, ou cerca de 1 em 8. Mas, como a maioria dos jogos de azar (e é exatamente o que são as loterias), esse cálculo é enganoso, e as chances reais são muito menores. Esse número de 1 em 8 seria verdadeiro se houvesse apenas seis bolas numeradas entre as 49, e tivéssemos de acertar apenas uma das seis. O que nos pedem é muito mais difícil: acertar seis bolas em 49, todas elas numeradas.

As chances são realmente muito pequenas: perto de 1 em 14 milhões. Por que tão pequenas? Porque nossa chance de acertar o primeiro número é de 1 em 49, a chance de acertar o segundo entre os 48 restantes na máquina é de 1 em 48; para o terceiro, é de 1 em 47; e assim por diante, até chegar ao sexto número, que é de 1 entre as 44 bolas restantes. Como a chance de qualquer bola específica sair da máquina é aleatória, e, portanto, independe das chances das outras bolas, a probabilidade de adivinhar corretamente todos os seis números de qualquer conjunto dado é calculada multiplicando-se todas essas probabilidades – $(1/49) \times (1/48) \times (1/47) \times (1/46) \times (1/45) \times (1/44)$ –, o que resulta quase exatamente em 1 em 10 bilhões. Os organizadores das loterias nos facilitam um pouco a vida ao não exigir que acertemos também a ordem exata em que os números saem da máquina. Aceitam qualquer uma das 720 ordens diferentes dessas seis bolas (digamos, 2, 5, 11, 34, 41, 44 ou 34, 2, 5, 11, 44, 41 etc.). Então, a chance de acertarmos os mesmos números são de mais ou menos 1 em 10 bilhões vezes 720, o que dá aproximadamente 1 em 14 milhões. Só para o caso de você achar que a chance não é tão ruim, imagine o seguinte: é como se os organizadores da loteria fizessem no chão uma pilha de dez pacotes de açúcar

de 1 quilo e pedissem que você cate na pilha o único grão pintado de preto – numa única catada, e de olhos vendados. Boa sorte.

Assim, as chances são de que jamais ganhemos o grande prêmio, mesmo jogando pelo resto da vida. Realmente, pode-se mostrar que o apostador médio de loteria no Reino Unido tem uma chance maior de cair morto durante a meia hora que leva para assistir ao sorteio e dar o telefonema reclamando o prêmio. Contudo, à espreita no meio desses números que nos decepcionam rotineiramente semana após semana, há uma importante lição de aleatoriedade. De fato, ela é tão importante que merece ser elevada ao status de lei da ausência de leis.

TERCEIRA LEI DA AUSÊNCIA DE LEIS

A verdadeira aleatoriedade não tem causa ou motivo, e, em última análise, é desprovida de padrões. Mas isso não significa que não tenha todos os padrões em toda escala. De fato, nas escalas em que a encontramos, a aleatoriedade é chocantemente propensa a produzir regularidades que seduzem a nossa mente ávida de padrões.

A evidência para essa lei pode ser encontrada assistindo-se regularmente aos sorteios de loteria na TV – ou, para aqueles que necessitam de uma gratificação mais rápida, conferindo os arquivos on-line de resultados anteriores. O exame ao acaso (de que outro jeito?) dos seis números ganhadores da loteria nacional do Reino Unido ao longo de algumas semanas não revelará qualquer padrão óbvio – aparentemente confirmando a nossa crença de que a aleatoriedade de fato significa ausência de padrões em toda escala. Por exemplo, eis os oito conjuntos ganhadores no sorteio do Reino Unido em junho de 2014:

14, 19, 30, 31, 47, 48

5, 10, 16, 23, 31, 44

11, 13, 14, 28, 40, 42

9, 18, 22, 23, 29, 33

10, 11, 18, 23, 26, 37

3, 7, 13, 17, 27, 40

5, 15, 19, 25, 34, 36

8, 12, 28, 30, 43, 39

À primeira vista, parecem 48 números sem nenhum padrão, viés ou sequência óbvia, exatamente como seria de esperar. Mas olhe de novo, dessa vez procurando o padrão mais básico possível em bolas de loteria: dois números consecutivos. Quatro dos oito conjuntos contêm essa “sequência”; de fato, o primeiro conjunto apresenta duas delas. É provável que você não as tenha percebido porque são padrões tão triviais que se esquivam até da renomada capacidade do *H. sapiens* para identificar padrões. No entanto, essa é uma insinuação dos padrões que a aleatoriedade pode nos mostrar e como eles seguem certas leis – tudo em aparente desafio às nossas crenças sobre a aleatoriedade. Usando um astucioso ramo da matemática chamado análise combinatória, é possível contar as maneiras de obter sequências de diferentes comprimentos entre os seis números, e descobre-se que se deve esperar *pelo menos* dois números consecutivos em metade de todos os sorteios de loteria tipo “6 em 49”. Logo, nos oito sorteios durante junho de 2014, deveríamos esperar que cerca de quatro tivessem uma sequência de dois ou mais números, e foi exatamente isso que obtivemos – e que obteríamos na maioria dos meses, se nos déssemos ao trabalho de checar.

Antes que alguém pense que isso poderia ajudar a predizer quais números vão ganhar a cada semana, não se esqueça de que ainda não temos ideia de *quais* dois ou mais números estarão em sequência: isso é aleatório, e portanto imprevisível. O que mostramos é que acontecerá em *algum* par ou sequência mais longa de números. Mesmo assim, o exemplo encerra algumas lições importantes para nós acerca dos padrões de aleatoriedade. Primeiro, mostra que os padrões não apenas são possíveis na aleatoriedade, na verdade, eles são surpreendentemente comuns – e a proporção em que aparecem pode ser calculada. Segundo, ressalta o fato de que muitas amostras de aleatoriedade –

incluindo sorteios de loteria – têm montes de padrões, mas nós deixamos de percebê-los porque os consideramos “insignificantes”; em outras palavras, devemos ter cautela ao tentar enxergar padrões “significantes” na aleatoriedade, porque os padrões estão no olho do observador. Terceiro, se, por um lado, ao sermos muito *específicos* em relação ao que queremos da aleatoriedade, reduzimos as chances de obter o desejado (por exemplo, o conjunto de seis bolas vencedoras do grande prêmio), quando somos muito *vagos* (por exemplo, “qualquer par consecutivo”), a chance de obtê-lo aumenta grandemente.

Podemos pôr tudo isso em funcionamento procurando outros padrões nessas amostras de aleatoriedade que vemos nos sorteios da loteria. Os espectadores do sorteio 1 310 da loteria nacional do Reino Unido, em 12 de julho de 2008, ficaram atônitos ao presenciar nada menos que quatro números consecutivos entre as seis bolas tiradas das 49 na máquina randomizadora: 27, 28, 29, 30. Um mês depois, a máquina da loteria despejou outros padrões, dessa vez de três números consecutivos entre os seis: 5, 9, 10, 11, 23, 26. Mesmo sendo mais impressionantes que meros pares, esses padrões ainda são surpreendentemente comuns – no mínimo porque não nos preocupamos em saber quais sequências de três ou quatro bolas formam o padrão. Cálculos combinatórios mostram que mesmo a surpreendente sequência de quatro números consecutivos deveria surgir em média uma vez em cada 350 sorteios – então, indiscutivelmente, a maior surpresa é porque foram necessários 1 300 sorteios para vê-la pela primeira vez (e, com toda a certeza, tem havido várias desde então).

À luz dessas percepções, o aparecimento de uma sequência completa de cinco números consecutivos no sorteio da loteria Fantasy 5 na Flórida, em 21 de março de 2011, não deveria parecer tão chocante. Mais uma vez, não estamos exigindo um conjunto específico de números, e isso faz com que seja mais fácil obtê-los. Realmente, é fácil fazer a síntese para ver isso. Seguindo o raciocínio para a loteria do Reino Unido, extrair cinco bolas de 36 no sorteio Fantasy 5 na Flórida numa sequência certa é possível em cerca de 45 milhões de maneiras. De novo, os organizadores nos facilitam a vida, e qualquer um dos 120 ordenamentos diferentes de cinco bolas é aceitável como vencedor, então há 375

mil maneiras de acertar as bolas premiadas. Mas, destes, apenas alguns serão totalmente consecutivos: o primeiro conjunto é $\{1, 2, 3, 4, 5\}$, depois $\{2, 3, 4, 5, 6\}$, e assim por diante até $\{32, 33, 34, 35, 36\}$. Há somente 32 conjuntos consecutivos, então a probabilidade de cinco números serem consecutivos é de $32/375\,000 = 1$ em 12 000. Como são realizados sorteios sete dias por semana durante o ano todo, isso significa que devemos esperar um intervalo aproximado de trinta anos entre cada exemplo de cinco bolas consecutivas. Dê tempo suficiente para a aleatoriedade, e ela vai acabar surgindo com alguma coisa. Nesse caso, a primeira apareceu depois de 23 anos, o que é um pouquinho cedo, mas não escandalosamente cedo.

Há mais uma lição valiosa sobre aleatoriedade que podemos aprender dos sorteios de loteria – e um estudo de caso surgiu no sorteio de meio de semana da loteria do Reino Unido pouco depois do conjunto de quatro bolas de numeração consecutiva. Primeiro veio uma trinca 9, 10, 11; depois outra, 32, 33, 34, na semana seguinte; e depois outra, 33, 34, 35, na semana depois da segunda.

Dessa vez temos um aglomerado de padrões. Então, o que podemos tirar daí? Nada, fora a surpreendente demonstração de como a verdadeira aleatoriedade pode aparecer nesses aglomerados. Os cálculos combinatórios mostram que, a longo prazo, essas trincas surgirão em 1 entre 26 sorteios desse tipo de loteria. Mas a aleatoriedade, com sua costumeira falta de causa ou motivo, não tem como se ater rigidamente a essa proporção. Às vezes as trincas serão largamente espaçadas, às vezes virão em aglomerados, como aconteceu em 2008. Apenas adeptos das teorias da conspiração tendem a enxergar alguma coisa nesses aglomerados. É algo bem diferente quando os padrões produzidos pela aleatoriedade representam não números de loteria, mas, digamos, casos de câncer numa cidade. Talvez haja algo nesses padrões, talvez não haja, mas mesmo aí devemos nos lembrar de que a aleatoriedade é capaz de produzir padrões e aglomerados de padrões com surpreendente facilidade.

Às vezes a loteria faz coisas que podem produzir sorrisos até entre os matemáticos. Depois de despejar padrões simples em julho e agosto, em 3 de setembro de 2008 a loteria do Reino Unido cuspiu seu padrão ainda mais

sofisticado: 3, 5, 7, 9, quatro números ímpares consecutivos. E depois disso voltou a fazer durante meses o que “se espera” que a aleatoriedade faça: ser enfadonha, chata e sem padrões.

Muitos matemáticos dizem que apostar na loteria é uma tremenda estupidez. Eles apontam as chances ridiculamente pequenas de ganhar o prêmio (lembra-se dos dez sacos de açúcar e do grão único?) e o fato de que os organizadores criam loterias para que os jogadores tenham de gastar mais que o prêmio médio em bilhetes para haver uma chance decente de ganhar. O que é verdade, embora se possa argumentar que pagar um bilhete aumenta infinitamente a chance de ganhar, de zero para 1 em 14 milhões, o que é um bocado. No entanto, como vimos, ainda que você tenha de estar “dentro dessa” para ganhar, algumas lições inestimáveis sobre aleatoriedade podem ser aprendidas de graça com qualquer loteria.

Conclusão

A maioria de nós acha que sabe como é a aleatoriedade: bacana, regular e totalmente carente de qualquer padrão ou aglomerado. A realidade é bem diferente – como atestam os números que surgem durante os sorteios de loteria. Eles revelam toda sorte de padrões e aglomerados. Mas embora a frequência desses padrões seja previsível, sua exata identidade nunca o é.

8. Aviso: há muito X por aí

EM MAIO DE 2014, foi registrado um veredito de suicídio para um jovem de dezesseis anos que se asfixiou num dormitório em Hale, Grande Manchester. William Menzies era um aluno com boas notas e sem nenhum problema óbvio. Mas o médico-legista notou algo que o deixou preocupado – algo que conectava a tragédia com outro caso de suicídio de adolescente com o qual lidara pessoalmente, com mais dois outros que encontrara. Todas as vítimas tinham se matado depois de jogar um videogame. E não um videogame qualquer, tampouco, mas o campeão de vendas *Call of Duty*, no qual os jogadores participam de ações em guerra virtuais.

Entre seus milhões de fãs – e críticos – *Call of Duty* é conhecido por seu imersivo realismo. O famoso terrorista solitário Anders Breivik alegou ter usado o jogo como treinamento antes de assassinar 77 pessoas na Noruega em um dia em julho de 2011. Será que *Call of Duty* é tão realista que deflagra os mesmos efeitos colaterais que os combates da vida real, como distúrbio de estresse pós-traumático, depressão e até pensamentos suicidas? O legista ficou preocupado o bastante com os riscos para emitir uma advertência, instando os pais a manter os filhos longe desses jogos.

Nem todo mundo ficou convencido com sua lógica. Entre os céticos estava o dr. Andrew Przybylski, psicólogo experimental no Oxford Internet Institute. Ele destacou que milhões de adolescentes jogam *Call of Duty* no Reino Unido; assim, não deve ser surpresa que alguns deles se suicidem. O dr. Przybylski ressaltou seu argumento com uma analogia: montes de adolescentes usam jeans, o que tornava provável que muitos dos que se suicidam estivessem usando jeans naquele momento. Será que faz sentido concluir que os jeans provocam suicídio?

Assim enunciado, fica evidente por que esses argumentos realmente não vingam. Primeiro, eles focalizam apenas em parte o que é necessário para

estabelecer o caso de um vínculo causal entre X e Y. Isto é, focalizam a probabilidade surpreendentemente alta de que adolescentes que se suicidam tenham jogado *Call of Duty* pouco antes de sua morte. Mas como nós *sabemos* que é surpreendentemente alta? O único meio é colocar a situação no contexto – o que significa compará-la com a probabilidade de adolescentes que *não* se suicidam terem jogado *Call of Duty* recentemente. E se estamos lidando com algo tão ubíquo como adolescentes jogando *Call of Duty*, você pode apostar que uma alta proporção de adolescentes perfeitamente felizes também terá jogado.

O exemplo destaca uma circunstância geral: tome cuidado ao acreditar que X explica Y se X for algo muito comum. Mas o inverso também vale: se algum efeito é muito comum, tenha cuidado ao jogar a culpa do seu surgimento em alguma causa específica – e, se ele for muito comum, é provável que tenha múltiplas causas. Um exemplo clássico ocupou há pouco tempo as manchetes, dizendo respeito a um importante debate sobre saúde pública no Reino Unido. Estatinas são drogas que reduzem o colesterol e vem se demonstrando que diminuem a possibilidade de morte entre pessoas com risco relativamente elevado de doença cardíaca. Isso levou alguns médicos especialistas a sugerir que mesmo pessoas com pouco ou nenhum risco extra também devem tomar estatinas como medida preventiva. A proposta provocou uma briga enorme tanto entre especialistas quanto entre pacientes. Alguns a veem como um passo no sentido da “medicalização” de todos, pela qual nós engolimos pílulas em vez de levar uma vida mais saudável. Contudo, a maior parte da preocupação gira em torno de difundidos relatórios de fadiga e dores musculares entre os que tomam estatinas. Ninguém está desprezando o desconforto que esses sintomas representam – embora alguns digam que é um preço pequeno a pagar em troca da redução da possibilidade de morte prematura. O que ninguém pôde questionar, no entanto, foi o fato de que esses sintomas são extremamente disseminados na população em geral. E isso leva à suspeita de que o elo com as estatinas talvez seja inteiramente ilegítimo.

Essa possibilidade foi posta à prova pela análise de estudos envolvendo coletivamente mais de 80 mil pacientes.¹ Essas pesquisas eram do tipo “duplo-

cego”: nem os pacientes *nem* os pesquisadores sabiam quem recebia estatinas e quem recebia um placebo. Os dados mostraram que cerca de 3% das pessoas que tomavam estatinas de fato sofriam de fadiga, e impressionantes 8%, de dores musculares. Tudo muito preocupante, até se descobrir que proporções praticamente idênticas dos pacientes que recebiam placebo também apresentavam os mesmos sintomas. Em outras palavras, não há razão para pensar que ingerir estatinas aumenta o risco de desenvolver seus efeitos colaterais mais “conhecidos”. Eles são tão comuns que há uma chance relativamente alta de que alguém que comece a tomar estatinas também experimente um surto de fadiga ou dores – e, de forma absolutamente compreensível, jogue a culpa nas drogas.

Compreensível, talvez – mas justificável apenas quando se exclui o risco de confundir ubiquidade com causalidade. E às vezes, para isso, são necessários estudos científicos completos envolvendo enormes quantidades de dados.

De modo estranho, toda uma classe de estudos científicos vem se justificando com base nesse tipo de raciocínio furado. Isso diz respeito ao aspecto talvez mais controverso da ciência experimental: o uso de animais. É inquestionável que experimentos em animais têm sido importantes em muitas áreas da medicina, desde a cirurgia até a pesquisa sobre o câncer. Tampouco se pode duvidar de que o uso de animais provoca fortes reações tanto dos setores pró quanto antivivissecção. O debate daí resultante é agressivo, até violento, e cada lado apresenta argumentos e contra-argumentos. Mas, para aqueles que apoiam o uso de animais, um argumento adquiriu poder quase talismânico: “virtualmente toda” conquista da medicina no último século dependeu de algum modo da pesquisa com animais.

Apesar de citada por pesquisadores famosos e mesmo pela Royal Society, a principal academia científica britânica, a justificativa para essa afirmação está longe de ser evidente. O argumento provém de um artigo anônimo num informativo que circulou pela Sociedade Americana de Fisiologia cerca de vinte anos atrás, e que não traz uma só referência para respaldar a impressionante afirmação. Mesmo assim, a conclusão é clara: se os cientistas quiserem

encontrar drogas capazes de salvar vidas, é vital continuar com os experimentos em animais. Entretanto, como o suposto elo entre suicídio e videogames, um aspecto-chave é negligenciado: a pura ubiquidade de experimentos em animais. Desde a tragédia da talidomida, na década de 1950, introduziu-se uma exigência legal para que toda droga nova passe por testes com animais antes de se permitir que seja testada em voluntários humanos, e muito menos liberada para o mercado. Como consequência, toda droga – independentemente de funcionar em seres humanos ou não – deve ser testada em animais. O fato de que todas as drogas bem-sucedidas tenham sido testadas em animais é um mero truísmo, e nada nos diz sobre o elo causal entre o uso de animais e o progresso da medicina. Dizer que significa algo faz tanto sentido quanto alegar que a prática igualmente ubíqua de vestir jaleco no laboratório é crucial para o progresso da medicina.

Como tal, a afirmação endossada pela Royal Society (entre muitas outras) é essencialmente vazia. No entanto, importa ressaltar que isso não quer dizer que os experimentos em animais não façam sentido. Significa que os cientistas precisam de evidências fortes se quiserem provar o valor de experimentos em animais. De modo surpreendente, pouco trabalho tem sido feito nessa área; o que foi feito é largamente inadequado para o propósito.² A evidência aponta para uma visão bem mais matizada desses experimentos do que qualquer um dos lados no debate está disposto a admitir. Sugere que modelos animais têm algum valor para detectar a toxicidade antes de realizar testes em seres humanos, mas são indicadores pobres em termos de segurança. Falando de forma mais prosaica, se Totó reagir mal a algum composto, é provável que os homens também reajam. Mas se Totó suportar bem o composto, isso nos diz muito pouco sobre o que nos acontecerá.

Conclusão

Provar que uma coisa causa alguma outra muitas vezes é algo traiçoeiro – e carregado de perigos se a suposta causa ou o efeito for muito comum. Mostrar que a suposta causa sempre precede o efeito é um começo, mas, em tais casos, raramente suficiente.

9. Por que o espetacular tantas vezes vira “mais ou menos”

NÓS VEMOS ISSO por todo lado, desde filmes de sucesso estrondoso, cujas sequências são sofríveis, até ações da bolsa que sobem às alturas e de repente despencam. Os espetaculares rojões de hoje têm o hábito de virar os estalinhos molhados de amanhã. Especialmente irritante é a maneira como perdem a magia no exato instante em que os notamos. Nossos amigos nos falam de um restaurante absolutamente *espetacular* onde jantaram na semana anterior, então resolvemos experimentar – e ele é só mais ou menos. Apostamos numa jogadora de tênis que ocupa as manchetes pelas performances estelares – só para vê-la afundar de volta na manada das jogadoras medíocres. Às vezes é difícil não pensar que tudo é só propaganda, e que a maioria das coisas está, bom, apenas na média. O caso é que, quando se trata de entender esse fastidioso equívoco da vida, você está no caminho certo.

Todo mundo já ouviu a frase “Não acredite em propaganda”, o que certamente nenhum de nós faria se pudéssemos distingui-la das opiniões confiáveis. A propaganda geralmente é tomada como símbolo de algum tipo de exagero da verdade, mas isso pressupõe que saibamos realmente qual a verdade. É aqui que saber um pouquinho de probabilidade talvez seja útil. Primeiro, a lei das médias nos diz que, quando tentamos avaliar o desempenho típico de qualquer coisa que possa ser afetada por efeitos aleatórios, devemos coletar uma profusão de dados. Claramente, faz pouco sentido esperar uma sequência espetacular de um autor de primeira viagem ou de um diretor de cinema principiante, pois ambos nos deram apenas um ponto na curva para julgá-los.

Mas a teoria da probabilidade também nos adverte de que coletar montes de dados não basta; eles também precisam ser *representativos*. Por definição, apenas os dados sobre desempenhos excepcionais não são representativos.

Contudo, é exatamente isso com que somos alimentados quando lemos críticas empolgadas, vemos frases de efeito em cartazes ou ouvimos peritos em economia delirando sobre algum novo investimento com o valor nas alturas. Por conseguinte, quando chega a hora de avaliar eventos excepcionais, devemos sempre *temer o fenomenal*. Basear nosso julgamento apenas em evidências de desempenhos excepcionais nos torna propensos a cair no traiçoeiro efeito conhecido como regressão à média. Identificada pela primeira vez há quase 150 anos pelo polímata inglês sir Francis Galton, ela ainda não é tão conhecida quanto deveria, apesar de sua onipresença.

Talvez as vítimas mais comuns da regressão à média sejam os torcedores esportivos. Eles já a viram em ação inúmeras vezes, e podem muito bem ter desconfiado de que alguma coisa estranha estava acontecendo – mas raramente identificaram o quê. A coisa funciona assim: no começo do campeonato, tudo parece correr como sempre – seu time ganha algumas partidas, perde outras. Aí ele desembesta e começa a cair na tabela até a zona de rebaixamento. Alguma ação se faz necessária; cabeças precisam rolar. Depois de uma sequência de derrotas, o time entende o recado e demite o técnico. Com total certeza a jogada dá certo: o time começa a ter uma atuação melhor com o novo técnico e as novas táticas. Mais aí tudo passa a dar errado de novo. Depois de uma sequência de atuações sólidas, o time começa a escorregar. Alguns meses apenas depois da revolução, ele parece estar exatamente na mesma – e recomeça o falatório sobre arranjar um novo técnico.

Isso soa familiar mesmo para aqueles que não entendem nada de futebol. É porque o mesmo fenômeno pode ser observado em toda parte, desde escolas com baixo desempenho até o mercado de ações. A ideia básica por trás da regressão à média não é difícil de entender. A performance de um time – ou de uma escola, ou do preço das ações – depende de uma série de fatores, alguns óbvios, alguns nem tanto, mas todos contribuindo para o nível da “média”. Contudo, num dado momento, a performance real provavelmente não estará cravada na média. Em geral estará um pouco acima ou abaixo dela, como resultado apenas de uma variação aleatória. Essa variação pode ser surpreendentemente grande e persistir

por um longo tempo, mas no fim seus impactos positivos e negativos se equilibram, e a performance “regredirá” ao valor médio. O problema é que a regressão à média é especialmente forte nos eventos mais extremos, e estes são os menos representativos de todos. Quem agir somente com base nesses eventos extremos arrisca-se a ser vítima da parte mais cruel da regressão à média: sua capacidade de fazer com que uma decisão ruim pareça, de início, uma decisão boa.

Por exemplo, um técnico trazido para dirigir um time depois que este apresentou evidência “incontestável” de mau desempenho pode se beneficiar de uma sequência de bons resultados. No entanto, talvez a melhora não passe de uma regressão à média, quando o time retorna a seu nível de atuação típico após uma sequência aleatória ruim, que custou ao técnico anterior o emprego. Espere tempo suficiente, e o nível de atuação típico irá se reafirmar. No início, parece haver uma atuação brilhante dos jogadores sob as ordens do novo técnico; mas isso talvez represente uma sequência de sorte coincidente com a chegada do técnico; depois, os jogadores irão regressar à média – começando a parecer cada vez mais medíocres à medida que o tempo passa. Dessa forma, a aparente explosão desfrutada pelo time também começa a desaparecer. Claro que às vezes um time tem atuações ruins porque o técnico perdeu a mão. Mesmo assim, pesquisas feitas por estatísticos e economistas usando dados da vida real mostram que a regressão à média pode afetar, e afeta, os times, resultando na demissão e contratação de técnicos, mas com pouco efeito sobre o desempenho geral da equipe.

Uma vez conhecendo a regressão à média, você começará a vê-la em todo lugar. Isso acontece porque com frequência nos concentramos nos extremos. Tomemos as técnicas de gerenciamento destinadas a melhorar o desempenho. Muitos gerentes de produção estão convencidos de que o melhor motivador é o medo – e chegam a argumentar que possuem fortes evidências para provar isso. Toda vez que sua equipe tem uma performance seriamente inferior, eles a chamam para dar uma bronca – e a performance melhora. E não me venha com essa baboseira de recompensar a performance, diz o gerente entusiasmado: isso é

uma “óbvia” bobagem. Afinal, quando se concedem prêmios trimestrais para a equipe campeã de vendas, em geral ela fica “mais ou menos” no trimestre seguinte; e isso “claramente” é complacência.

OS IMPRESSIONANTES PODERES DE CURA DA REGRESSÃO À MÉDIA

Na busca de novas terapias, pesquisadores da área médica correm o risco de se enganar com a regressão à média, julgando ter encontrado uma cura milagrosa. Pela sua própria natureza, a procura desses tratamentos muitas vezes se concentra em pacientes com características anormais, como pressão sanguínea muito alta. No entanto, as anormalidades talvez não sejam mais significativas que desvios aleatórios da normalidade – que irão desaparecer com o tempo. Identificar esses efeitos é um desafio para os pesquisadores que testam uma droga nova, pois eles correm o risco de pensar que a substância provocou alguma melhora com o tempo, quando o estado do paciente simplesmente regressou à média. Eles lidam com o fato estabelecendo as chamadas experiências aleatoriamente controladas, nas quais pacientes são alocados de forma aleatória para receber ou não a droga, ou receber um inócuo “controle” placebo. Como ambos os grupos têm igual probabilidade de experimentar a regressão à média, seus efeitos podem ser anulados comparando-se as taxas relativas de cura nos dois grupos. Infelizmente, não há salvaguardas desse tipo quando um amigo nos recomenda um remédio para, digamos, dor nas costas. Carecendo de qualquer grupo de comparação, é difícil ter certeza de que qualquer benefício que eventualmente obtemos não seja apenas uma regressão à média. Certos médicos argumentam que pacientes que se acreditam curados pela “medicina alternativa”, como a homeopatia, melhoraram somente pela regressão à média. Os advogados dos homeopatas insistem, porém, em que foram realizados estudos levando essa possibilidade em consideração, e que eles demonstraram o evidente benefício do tratamento.

É verdade, os dados de performance parecem provar esse fato – a não ser que você conheça a regressão à média. O problema é que os chefes muito entusiasmados não aceitam bem a sugestão de que sua “prova inquestionável” para o aumento de eficiência não passa de um efeito estatístico – e essa deve ser outra razão para que tão poucos saibam sobre o assunto.

No entanto, nós devemos ao menos nos proteger da autoilusão. Por exemplo, quando se trata de fazer investimentos, precisamos ter muita cautela com ações exuberantes reverenciadas pelos sacerdotes financeiros. De hábito, eles

focalizam as performances fenomenais, dignas de manchetes – o terreno fértil clássico para a regressão à média. Mais uma vez, esse não é um risco teórico. O dr. Burton Malkiel, economista da Universidade de Princeton e flagelo de Wall Street, fez um estudo do que acontece com aqueles que investem em ganhadores “óbvios”.¹ Ele compilou uma lista dos fundos de ações com melhor desempenho no período de 1990 a 1994. Os primeiros vinte desses fundos estiveram acima do índice S&P 500 por uma impressionante margem média anual de 9,5%, e eram ganhadores “óbvios”. Malkiel então examinou como esses mesmos fundos se saíram durante os cinco anos seguintes. Coletivamente, tiveram um desempenho médio inferior, em mais que 2%, ao mercado de ações como um todo. O ranking dos três primeiros despencou de 1º para 129º, de 2º para 134º e de 3º para um desastroso 261º. Tamanho é o poder da regressão à média para nos ensinar lições de humildade.

Como os técnicos de futebol, porém, um punhado de gerentes de investimentos realmente parece saber o que está fazendo e consegue desempenhos impressionantes, que não podem ser desprezados como casualidade estatística. Um deles é a ex-lenda de Wall Street Peter Lynch, cujo Magellan Fund previu um desempenho estarrecedor durante as décadas de 1970 e 1980. Infelizmente, a evidência sugere que a maioria dos gerentes “estrelas” se beneficia apenas temporariamente da regressão à média, e estão destinados a sumir após alguns anos – e levar nossos investimentos com eles.

Conclusão

Quando se trata de tomar decisões baseadas em desempenho, tenha medo do fenomenal. Um desempenho excepcional, por definição, é tudo, menos representativo. E isso aumenta especialmente sua probabilidade de decepcionar, cortesia da Grande Equilibradora que é a regressão à média.

10. Se você não sabe, vá pelo aleatório

DURANTE UMA COLETIVA DE IMPRENSA em fevereiro de 2002, o então secretário de Defesa dos Estados Unidos, Donald Rumsfeld, foi indagado sobre o risco de o ditador iraquiano Saddam Hussein fornecer a terroristas armas de destruição em massa. Claramente irritado com a pergunta, Rumsfeld deu uma resposta que ficou famosa:

[Como] sabemos, há conhecidos conhecidos; há coisas que sabemos que sabemos. Nós também sabemos que há desconhecidos conhecidos; isso quer dizer que sabemos que há coisas que não sabemos. Mas há também desconhecidos desconhecidos – coisas que não sabemos que não sabemos.^{1b}

Essa foi uma resposta que provocou choque e estupefação entre os críticos de Rumsfeld. Alguns a tomaram como prova positiva de que o Pentágono estava sob controle de um lunático. Outros a encararam como simplesmente risível: a Plain Speaking Society, Sociedade pela Clareza da Fala, do Reino Unido, concedeu a Rumsfeld um prêmio especial pelo absurdo. Alguns, porém, viram sua resposta como a declaração sucinta de uma perturbadora verdade acerca da confiabilidade do conhecimento: há ignorância, e depois há a ignorância da ignorância. Nada podemos fazer em relação a esta última – pois como podemos nos proteger de algo que nem sequer sabemos que existe? Na verdade, *existe* algo que podemos fazer para no mínimo reduzir a ameaça dos desconhecidos desconhecidos. E, o que é ainda mais surpreendente, este algo está na aleatoriedade.

Com sua proverbial falta de causa ou motivo, a aleatoriedade parece uma fonte estranha de segurança na busca de conhecimento. Contudo, é exatamente por isso que ela se torna tão valiosa: a aleatoriedade incorpora a liberdade a partir de premissas subjacentes, que é onde nossa ignorância às vezes se manifesta da maneira mais destrutiva. Essa potente característica chamou a atenção de cientistas sobretudo pelos esforços de um dos fundadores da

estatística moderna, cujo nome aparecerá diversas vezes ao longo deste livro: Ronald Aylmer Fisher. Depois de se graduar em matemática na Universidade de Cambridge, mais ou menos um século atrás, Fisher ficou fascinado com o desafio de extrair dos dados as informações mais confiáveis – especialmente nas complexas e complicadas ciências da vida. Trabalhando como estatístico num laboratório de pesquisa em agricultura, ele concebeu uma série de técnicas para extrair informações de experimentos assolados pelos desconhecidos desconhecidos que empestiavam esse tipo de pesquisa – por exemplo, a variabilidade da fertilidade do solo. Seu livro-texto sobre a análise dos resultados, *Statistical Methods for Research Workers*, publicado em 1925, tornou-se talvez o livro de estatística mais influente já publicado. Mas a principal ferramenta por ele recomendada era a “aleatorização”, que, segundo declarou Fisher, “ libera o experimentador da ansiedade de considerar e estimar a magnitude das inumeráveis causas pelas quais seus dados podem ser perturbados”.²

Em nenhum outro lugar seu conselho foi usado de maneira mais adequada que na medicina, área onde se provou vital na busca de terapias efetivas. Já no século XIV, o poeta e estudioso italiano Petrarca falava em testar novas poções arranjando “centenas ou um milhar de homens” com características idênticas, tratando só metade deles e observando como reagiam em comparação aos que não haviam sido tratados.³ Como nos demais aspectos os homens eram iguais, qualquer diferença provavelmente resultaria do tratamento. Tudo muito simples, exceto uma coisa: o que queremos dizer com pessoas “idênticas”? Em tese, elas precisam ser idênticas ao paciente típico destinado a receber tratamento se passar pela inspeção. O problema é que as pessoas têm naturalmente um monte de diferenças: físicas, emocionais e genéticas, entre outras. O impacto dessas diferenças sobre o resultado cria uma porção de “desconhecidos conhecidos”. Acrescentem-se a eles os desconhecidos desconhecidos, e o método descrito por Petrarca começa a parecer simplista demais.

É aí que entra a aleatoriedade para salvar a situação. Em vez de tentar abarcar tudo que possa afetar a maneira como as pessoas reagem (e, quase com

certeza, fracassar), pegamos uma amostra dos pacientes e aleatoriamente os alocamos para receber a nova terapia ou ficar sem tratamento (ou receber placebo). Sendo uma amostra, ela jamais será perfeita, mas claramente seria tão boa de usar quanto a maior amostra possível. O próprio Petrarca mencionou isso – mas não a característica adicional de aleatoriedade recomendada por Fisher. Ao alocar pacientes totalmente ao acaso, reduzimos o risco de que a amostra seja “enviesada”, por acidente ou outros motivos, para o lado daqueles que poderiam (ou não) beneficiar-se do tratamento.

Tendo usado a aleatoriedade para resolver o problema de pacientes “idênticos”, podemos pôr em ação o restante da sugestão de Petrarca: criar dois grupos de pacientes, o grupo de tratamento, aqueles que recebem a terapia, e o grupo de controle, dos que recebem alguma terapia comparativa (ou talvez apenas um placebo). É inteiramente possível que um dos grupos tenha mais pacientes com, digamos, algum traço genético desconhecido que atrapalhe o tratamento. Mas, ao usar muitos pacientes escolhidos ao acaso, há uma boa chance de termos números bastante similares desse tipo de paciente em ambos os grupos. Com o viés que podemos introduzir assim mitigado, a avaliação da terapia torna-se mais confiável.

Esse, porém, não é o único benefício de se empregar a aleatorização. Uma vez de posse dos resultados, eles precisam ser interpretados corretamente. Por exemplo, se surgir uma diferença entre grupos de pacientes, sugerindo que a terapia é efetiva, sempre é possível que aquele seja apenas um resultado casual. Por outro lado, o fracasso em achar uma diferença pode ser resultado de se usar um número pequeno demais de pacientes. A quantificação das chances desses resultados necessita da teoria da probabilidade, e isso será mais simples e digno de confiança se pudermos assumir que não há vieses em ação. A aleatorização terá sucesso aí – e até ajuda a lidar com algumas questões éticas traiçoeiras. Pesquisadores inescrupulosos podem querer ministrar suas drogas a pacientes menos enfermos, enquanto outros recebem uma terapia antiga menos efetiva – ampliando assim as chances de a nova droga dar bons resultados. Por sua vez, pesquisadores compassivos podem querer dar a nova terapia a pacientes que em

outras circunstâncias teriam pouca esperança... mas isso significaria condenar outros pacientes a receber o tratamento menos efetivo, se os pesquisadores fossem bons juízes da provável efetividade das suas terapias. Contudo, eles não são: uma análise feita em 2008, de mais de seiscentos experimentos controlados aleatoriamente de tratamentos de câncer – tratamentos considerados dignos de serem testados em pacientes pelo Instituto Nacional do Câncer dos Estados Unidos desde meados dos anos 1950 –, descobriu que apenas 25 a 50% se mostraram bem-sucedidos.⁴

Esses dilemas éticos são evitados simplesmente insistindo na alocação aleatória de cada grupo – e, mais ainda, por alguém de fora, não relacionado com o experimento.

Em 1947, o Conselho de Pesquisa Médica do Reino Unido decidiu testar o poder da aleatoriedade num estudo pioneiro acerca da eficácia do antibiótico estreptomicina contra a tuberculose. Não foi um teste muito grande: cerca de cem pacientes foram alocados aleatoriamente para receber ou tratamento-padrão de simplesmente ficar de cama ou repousar e tomar o antibiótico. Para evitar que médicos ou pacientes gerassem viés no resultado sabendo quem estava recebendo o quê, todos foram mantidos no escuro (“às cegas”) sobre o resultado do processo aleatório de seleção. Depois de seis meses, os resultados lá estavam – e eram aparentemente impressionantes: dos cinquenta e poucos pacientes que receberam antibiótico, a taxa de sobrevivência foi quase quatro vezes maior que a das contrapartes que tinham apenas ficado em repouso. Aquele era um experimento pequeno, todavia, os testes estatísticos sugeriram uma diferença tão grande que provavelmente não era casual. Hoje, esses chamados Estudos Clínicos Randomizados (ECRs) “cegos” tornaram-se o padrão-ouro na testagem da eficácia de novas terapias. Centenas de milhares vêm sendo realizados, alguns envolvendo dezenas de milhares de pacientes, e os resultados têm beneficiado a saúde de incontáveis milhões de pessoas. Tudo isso presta testemunho ao potencial da aleatoriedade para reduzir o impacto da ignorância – tanto conhecida quanto desconhecida. Seu sucesso na medicina tem estimulado

tentativas de usar o método ECR em outras áreas de pesquisa destinadas a atacar males como a pobreza e o crime juvenil (ver Box a seguir).

DANDO À POLÍTICA GOVERNAMENTAL O TRATAMENTO DA ALEATORIEDADE

O sucesso dos Estudos Clínicos Randomizados (ECRs) das drogas para determinar “o que funciona” provocou o interesse em usar a mesma ideia em outras áreas – como testar políticas governamentais. Os políticos têm a reputação de lançar grandes esquemas com base em pouco mais que palpites e situações ocasionais. Não seria melhor testar suas ideias usando a aleatorização, ou randomização, para combater suas premissas de onisciência?

Essa é uma ideia atraente – pelo menos para aqueles comprometidos com a noção de que a política pública deve se basear em fatos, e não em dogmas. Talvez seu maior sucesso até hoje tenha sido o programa de bem-estar social *Oportunidades*, no México, que combatia a pobreza dando dinheiro a famílias específicas em troca de comparecimento regular à escola, exames médicos e apoio nutricional.⁵ A ideia de oferecer dinheiro como material de troca pela participação foi desprezada pelos críticos, sendo considerada ingênua. Então o governo respondeu testando a proposta por meio de um ECR. Centenas de moradores de vilas foram randomizados para tomar parte ou para atuar como controle, e o impacto do programa foi monitorado. Dois anos depois, esse impacto foi avaliado – e a política considerada efetiva para aumentar tanto o bem-estar quanto as perspectivas de futuro dos que dela tomaram parte. Em 2002 o programa foi estendido também para comunidades urbanas, e tem mostrado tamanho sucesso que vem sendo copiado em outros lugares – inclusive na cidade de Nova York.

Nem toda ideia de inspiração política tem sido beneficiada pelo método ECR. Tomemos a política Scared Straight^c para lidar com delinquentes juvenis: o método consiste em fazer com que os jovens presenciem os horrores que os aguardam se acabarem na cadeia. Batizado a partir de um documentário americano de mesmo nome, produzido em 1978, sugeria que os indolentes se corrigiam depois de serem expostos à vida dos “sentenciados à perpétua” na cadeia de Nova Jersey. Alguns políticos reclamaram seu uso mais generalizado, porém, felizmente, nem todo mundo estava disposto a confundir fato ocasional e evidência. O esquema foi submetido a uma série de ECRs, e, quando analisados, em 2013, os resultados mostraram que a política na verdade era menos que inútil: os que tomaram parte dela apresentaram índices *mais altos* de delinquência do que os que foram deixados em paz.⁶

Felizmente, há sinais de que alguns governos estão começando a ver que ECRs são o meio mais seguro de descobrir “o que funciona”, em lugar das intuições.⁷

No entanto, apesar de todo o seu poder, o ECR não é um infalível guia para “o que funciona”, como alguns parecem pensar. Embora a aleatoriedade, a

princípio, possa lidar com qualquer desconhecido desconhecido, na realidade ela recai no problema de tantas pesquisas feitas *por* seres humanos *em* seres humanos. Por exemplo, é fácil randomizar pessoas uma vez que tenham sido recrutadas pelos experimentadores – mas e se os recrutadores somente engajarem determinados tipos de pessoa? Ao longo dos anos, estudos randomizados têm levado os psicólogos a um mundo de percepções sobre a natureza humana. Contudo, as exigências de custo, tempo e conveniência significam que muitos desses insights vieram de estudos randomizados de tipos humanos claramente não aleatórios: estudantes de psicologia americanos. Em 2010, pesquisadores da Universidade da Colúmbia Britânica, no Canadá, publicaram uma análise acerca de centenas de estudos publicados em preeminentes jornais e revistas de psicologia, e descobriram que mais de dois terços dos participantes nas pesquisas vinham dos Estados Unidos e, entre eles, dois terços eram graduados em psicologia. Pior ainda, os pesquisadores descobriram que esses estudantes são especialmente não representativos dos seres humanos “típicos” – em sua esmagadora maioria, vêm de sociedades ocidentais, educadas, industrializadas, ricas e democráticas.^{8d}

Os vieses também podem aparecer durante um ECR – por exemplo, quando somente certos tipos de pessoa se mostram capazes (ou dispostas) de seguir rigidamente uma dieta alimentar restrita. Quem sabe por que elas caem fora? Talvez por mera casualidade, talvez não; de todo modo, isso é capaz de solapar a “validade externa” dos resultados – ou seja, o quanto eles se aplicam a você ou a mim. A verdade é que há uma enorme quantidade de maneiras pelas quais tudo, desde drogas até suplementos nutricionais, pode funcionar direito em estudos científicos, mas fracassar no mundo real.⁹

E esses são apenas os ECRs dos quais ouvimos falar. Nem a aleatoriedade nos protege no chamado viés de publicação, no qual achados de pesquisas considerados inconclusivos, tediosos ou “inúteis” simplesmente nunca são publicados. Vários estudos vêm mostrando que os resultados positivos têm maior probabilidade de ser publicados que os negativos ou inúteis.¹⁰ As causas disso são tema de debates acalorados. Alguns culpam práticas negligentes por

parte dos pesquisadores; outros alegam que as publicações científicas são ávidas demais por descobertas estrondosas. Companhias farmacêuticas têm sido acusadas de enterrar resultados negativos para proteger o valor de suas ações. Indubitável é o efeito potencialmente pernicioso que o viés de publicação pode provocar sobre tentativas de responder a perguntas fundamentais juntando num mesmo saco toda a evidência publicada. A “meta-análise” resultante possui uma propensão otimista, com consequências potencialmente ameaçadoras para a vida do público.

Finalmente, há o problema dos pesquisadores espertalhões. A aleatoriedade é impotente para contra-atacar o viés introduzido por pesquisadores que montam um ECR especificamente para chegar à resposta “certa”. ECRs supervisionados por companhias farmacêuticas são criticados por usar modelos “espantalhos”, nos quais a nova droga é comparada a algum remédio inapropriadamente inócuo – aumentando assim as chances dos resultados espetaculares.¹¹

Como todas as criações humanas, o ECR pode ser subvertido de um sem-número de maneiras. Mas o uso que fazem da aleatoriedade assegura que, com todos seus defeitos, ainda é o melhor meio que temos para nos proteger do delírio da onisciência.

Conclusão

A própria ausência de leis na aleatoriedade torna-a inestimável para cortar fora premissas mal formuladas – tanto as conscientes quanto as inconscientes – e práticas questionáveis. No entanto, quando mal utilizada ou utilizada parcialmente, ela faz uma pesquisa de má qualidade assumir ares de “científica”.

^b Em inglês o texto se torna ainda mais engraçado, já que “conhecer” e “saber” são o mesmo verbo, *to know*. Para que o leitor possa ter ideia, aí vai o texto no original: “[As] we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don’t know we don’t know.” (N.T.)

^c Impossível aqui reproduzir o poder de síntese do inglês; o nome significa mais ou menos “Amedrontar para enquadrar”. (N.T.)

^d Essas cinco características em inglês – *Western, educated, industrialised, rich, democratic* – formam o acrônimo Weird, que significa “estranho”, “esquisito”, reforçando assim a ideia de seres humanos “atípicos”. (N.T.)

11. Nem sempre é ético fazer a coisa certa

PENSANDO EM TROCAR adoçantes artificiais por açúcar? Pense outra vez: isso pode aumentar seu risco de diabetes. Preocupado com a possibilidade de perder o emprego? Em breve você poderá ter asma para acrescentar às suas desgraças. Tomando remédio para dormir porque está preocupado com todas essas ameaças à sua saúde? Você pode aumentar substancialmente o risco de contrair Alzheimer.

A lista de ameaças à nossa saúde parece ficar cada vez mais comprida; esses últimos acréscimos ergueram suas assustadoras cabeças na mídia no decorrer de apenas um mês, em 2014.¹ Todavia, muitas vezes é difícil saber o que concluir dessas histórias. Muitas parecem se basear em pesquisas realizadas por cientistas de boa reputação, e são divulgadas em publicações científicas de respeito. Mas o fato de que a comprovação de qualquer uma dessas ameaças específicas à saúde tantas vezes oscile de um lado para outro não ajuda em nada. Alguns anos atrás, o café foi condenado por aumentar o risco de câncer no pâncreas. Esse risco sumiu, e agora parece que o café é bom no combate ao câncer no fígado.²

Decidir o que fazer com base apenas em artigos da mídia claramente não tem nenhum cabimento. Cabe fazer uma avaliação científica adequada – e qual a melhor forma de conduzi-la que mediante aquele padrão-ouro para a investigação médica, o Estudo Clínico Randomizado (ECR)?

Não vamos tão depressa: esse estudo exigiria uma amostra aleatória de voluntários e a exposição deliberada de metade deles a algum fato de risco desconhecido e potencialmente pernicioso. Isso suscita alguns aspectos ético-legais óbvios. Mas não são os únicos problemas dos ECRs. Ao mesmo tempo que seria fascinante saber, digamos, se as pessoas que se tornaram vegetarianas são mais saudáveis que aquelas que comem carne, vai ser duro recrutar milhares de pessoas e dizer a metade delas que não poderá comer carne pelo resto da vida.

Mesmo com todas as suas vantagens, o ECR simplesmente não pode ser usado para investigar algumas questões – embora muitas vezes elas estejam entre as mais interessantes de se pesquisar. Assim, em vez disso, os pesquisadores usam o chamado estudo observacional. Como o nome sugere, esse estudo envolve observar dois grupos de pessoas, comparando-as em busca de evidência para o efeito sob exame. O que não soa muito diferente de um ECR, exceto pela ausência de seu traço mais poderoso: a randomização. Impossibilitados de recorrer ao seu poder para lidar com desconhecidos (tanto conhecidos quanto não), os estudos observacionais tentam desenvolver uma abordagem diferente. Como veremos, ele não é fácil de aplicar; na verdade, a evidência sugere que raramente isso é feito com efetividade. E esse é um grande motivo para que tantas histórias na mídia sobre riscos de saúde pareçam oscilar de lá para cá. A maioria se baseia em resultados de estudos observacionais – que, com demasiada frequência, revelam suas deficiências como substitutos dos ECRs.

O tipo mais comum de estudo observacional, de um formato chamado “caso-controle”, é um meio mais rápido de investigar o possível elo entre alguma condição médica e um fato de risco suposto. Estudos de caso-controle têm gerado um enxame de artigos sobre saúde que ganharam as manchetes, como o alegado elo entre tomar remédio para dormir e desenvolver doença de Alzheimer. Montar um estudo desses envolve encontrar um monte de gente com determinada condição (os “casos”) e um grupo correspondente de pessoas comuns (os “controles”). Os dois grupos são então comparados. O que os pesquisadores procuram são sinais de que as pessoas afligidas pela condição também tendem a ser aquelas com maior exposição à suposta causa.

O problema mais óbvio é conseguir um “grupo correspondente”. Sem randomização, os pesquisadores são forçados a decidir que critérios usar para estabelecer a correspondência dos dois grupos. Inclua critérios demais, e em pouco tempo você esgotará controles para formar pares com os casos; inclua critérios de menos, e a comparação vira uma piada. Escolha os critérios e a correspondência errados, e é possível que o verdadeiro elo simplesmente

desapareça no meio do processo de combinação. Inclua o risco de viés na escolha de quem é selecionado para cada grupo em primeiro lugar, e a oportunidade para os resultados não confiáveis torna-se óbvia.

Apesar de todo o seu potencial para fracassar, porém, os estudos de caso-controle são muitas vezes o único meio ético de investigar preocupações relativas a supostos riscos de saúde – especialmente para doenças incomuns, em que, de outra forma, caberia observar enormes quantidades de pessoas para se chegar a conclusões confiáveis. E os estudos apresentam alguns casos de sucesso espetacular a seu favor. O mais famoso é a evidência de um elo entre o câncer de pulmão e o hábito de fumar, revelado por um estudo de caso-controle publicado em 1950 por dois dos mais celebrados nomes da estatística médica: Austin Bradford Hill e Richard Doll. Armados de mais de mil casos e controles, eles conseguiram levar em conta uma enorme quantidade de fatores potencialmente relevantes, de idade e sexo até classe social, formas de aquecimento doméstico e mesmo exposição a outros poluentes. As proporções relativas de fumantes e não fumantes entre os casos de câncer e os livres da enfermidade apontavam para um robusto aumento de risco de câncer pulmonar em decorrência do fumo. No entanto, Hill e Doll foram mais longe e mostraram que o risco crescia com o aumento do consumo – uma relação “dose-risco” decerto consistente com o fato de o fumo ser uma causa de câncer pulmonar. Entretanto, esta não é uma prova: sem a randomização para combater pelo menos alguns dos vieses, há um risco substancial de que algum “desconhecido desconhecido” fosse na realidade o responsável. E havia o problema de que ambos, casos e controles, tinham sido pacientes de hospitais – o que talvez não representasse a população geral.

Doll e Hill responderam montando outro meio amplamente utilizado para investigar efeitos sobre a saúde: o estudo prospectivo de coorte. Desta feita, em vez de olhar para trás, em direção ao que teria deflagrado o efeito, um estudo prospectivo acompanha uma população grande – “coorte” – de pessoas sem saber quem será afetado. Assim, o efeito do “desconhecido desconhecido” é enfrentado escolhendo-se uma coorte de pessoas semelhantes sob muitos aspectos – por exemplo, de mesmo sexo e mesmo background socioeconômico.

Elas irão diferir, porém, quanto a terem sido expostas ou não à causa suspeita dos efeitos investigados.

Os dois pesquisadores decidiram focalizar os médicos, e, no começo dos anos 1950, tinham conseguido recrutar uma coorte de mais de 34 mil homens e mais de 6 mil mulheres, divididos entre fumantes e não fumantes. Propuseram-se então a seguir o destino dos dois grupos num estudo que durou até 2001. O que veio a se tornar o British Doctors Study encontrou evidência convincente de que fumar aumentava o risco de câncer de pulmão em aproximadamente dez vezes, e pelo menos vinte vezes em fumantes pesados.

Esse inequívoco sucesso incentivou os pesquisadores a se voltar para estudos de caso-controle e prospectivos para abordar uma legião de outras questões relacionadas à saúde. Isso fez com que as investigações se tornassem um feliz terreno de caça para a mídia: ela sempre pode rebater a acusação de fomentar temores apontando para o fato de os projetos terem sido publicados por uma ou outra “prestigiosa” revista científica. Entre os próprios pesquisadores, porém, as limitações dos estudos observacionais estão causando crescente preocupação. Grande parte dela diz respeito ao aparente insucesso de tantas investigações observacionais para chegar a qualquer tipo de consenso. Os resultados de estudos de caso-controle em particular vêm se tornando conhecidos pelas reviravoltas nas conclusões, e muitas vezes pesquisas sucessivas fracassam em replicar os achados anteriores, ou os contradizem categoricamente. Uma revisão do uso dos projetos que buscavam vincular enfermidades a genes específicos descobriu que, em 166 desses vínculos investigados múltiplas vezes, mal chegava a 4% a quantidade dos que foram replicados de maneira consistente.³ Estudos prospectivos de coorte, de maneira geral, saíram-se melhor, porém, mesmo aqueles em aparência mais impressionantes malograram em produzir conclusões convincentes.

Tomemos o presente furor acerca das implicações para a saúde de se comer carne. Em 2009, um enorme estudo de coorte abrangendo meio milhão de americanos monitorados por mais de dez anos revelou um elo claro entre o consumo de carne vermelha e o risco de câncer, de doenças cardiovasculares e

longevidade reduzida. Então, em 2012, uma abrangente pesquisa japonesa revelou que esse perigo não era real, e em 2013 um enorme estudo europeu apareceu com um monte de resultados misturados.⁴ Se até as pesquisas observacionais gigantescas, realizadas por especialistas renomados, não conseguem chegar a conclusões consistentes, de que adianta realizá-las? Para ser justo, é possível que os *dois* estudos estejam corretos. Diferenças na composição da carne bovina e a maneira como ela é preparada e consumida (na verdade, até entre os que a preparam e consomem) permitiriam concluir que a carne americana é menos saudável, pelo menos para os americanos. Mais uma vez, isso ressalta o problema da generalização no qual mesmo os ECRs podem tropeçar: a forma como a investigação foi conduzida produz resultados que se aplicam somente em circunstâncias especiais, não aplicáveis genericamente.

Mesmo assim, esses estudos podem muito bem ter sido vítimas da falta da randomização que dá poder aos ECRs. Para lidar com o problema, os pesquisadores tentaram identificar e cancelar (“controlar”) o impacto do maior número possível de fatores potencialmente indutores de erros, como histórico de fumar e ingestão de álcool. Isso requer cortar e fatiar os dados da coorte em uma porção de subgrupos. E significa que muitos dos achados baseiam-se apenas numa minúscula fração do impressionante meio milhão de pessoas que compõem a coorte total. Mesmo então, era possível que os resultados das duas pesquisas ainda estivessem sujeitos a vieses sutis. Em 2011, dois pesquisadores do Instituto Nacional de Ciências Estatísticas dos Estados Unidos lançaram luz sobre os perigos de tentar imitar os benefícios dos ECRs: eles examinaram as justificativas dadas nos estudos observacionais que depois foram testados contra o “padrão-ouro” de um ECR. Das 52 explicações feitas em doze estudos observacionais identificados, o número confirmado pelo ECR posterior foi... zero.⁵

Assim, quando defrontados com um artigo sobre qualquer risco (ou benefício) para a saúde revelado por um estudo observacional, como devemos reagir? Os epidemiologistas – os que trabalham nessa área de pesquisa – com frequência aplicam algumas regras práticas para decidir se os resultados devem

ser levados a sério (ver Box a seguir). Isso, por sua vez, levou ao surgimento de uma espécie de “hierarquia”^e quando se trata de estudos observacionais. A forma mais baixa de vida epidemiológica são pequenos estudos de caso-controle que alegam ter descoberto alguma evidência de um pequeno vínculo, antes não suspeitado, entre algum risco à saúde e uma causa implausível.

O exemplo prototípico disso é o suposto elo entre campos eletromagnéticos e leucemia infantil, cujos indícios surgiram pela primeira vez no fim da década de 1970. No decorrer dos anos, o elo tem sido examinado em muitos estudos de caso-controle com centenas de participantes; quando combinadas, essas pesquisas sugeriam um aumento significativo do risco de leucemia entre crianças expostas a campos eletromagnéticos de aparelhos e linhas elétricas. Todavia, a aplicação de algumas regras epidemiológicas práticas põe essa perturbadora conclusão sob outra luz. Por exemplo, apesar dos números aparentemente impressionantes de participantes desses estudos, os aumentos de risco mais preocupantes vieram daqueles expostos aos campos eletromagnéticos mais elevados – o que envolvia apenas alguns casos e controles. Além disso, nunca foi explicado de forma plausível como exatamente os campos eletromagnéticos devem provocar leucemia – enquanto existe uma profusão de potenciais fontes de viés e fatores enganadores capazes de simular esse vínculo. Tudo sugere que a evidência para risco de câncer provocado por campos eletromagnéticos é bastante frágil – e com toda a certeza os resultados têm se invertido repetidas vezes. Uma revisão das evidências feita em 2007 por uma equipe dos Centros de Controle de Doenças dos Estados Unidos excluiu os campos eletromagnéticos de sua lista de fatores de risco ambiental significativos para leucemia.⁶

SÉRIO OU ESPÚRIO? DANDO SENTIDO ÀS MANCHETES SOBRE SAÚDE

Todo estudo observacional aspira a identificar uma ligação causal genuína entre algum efeito sobre a saúde e certa atividade, desde comer *junk food* até morar perto de um reator nuclear. Entretanto, tudo que esse tipo de pesquisa oferece é uma evidência mais ou menos convincente de algum elo potencial. Como diz o ditado, “Correlação não é causalidade”, e separar as duas coisas nem sempre é algo fácil e direto. Existem, contudo, algumas regras

práticas que podem ser usadas para decidir que estudos levar a sério e quais merecem apenas um “Está bom, seja lá o que for”.

As mais úteis dessas regras foram sugeridas em meados da década de 1960 pelo professor sir Austin Bradford Hill, da Universidade de Londres, cujo estudo observacional sobre fumantes, iniciado nos anos 1950, estabeleceu um parâmetro raramente equiparado desde então.⁷ Inspirada nos critérios de Hill, eis uma lista proveitosa do que deve ser buscado:

Qual o tipo de estudo observacional? Um estudo “caso-controle”? Estes geralmente se debatem mais com o problema do viés que as pesquisas “prospectivas de coorte”.

Quão surpreendente é o achado? Seja especialmente cético em relação a argumentações que “vêm sem mais nem menos do nada”, acerca de efeitos antes desconhecidos sobre a saúde – em especial se a ligação for biologicamente implausível.

Qual o tamanho do estudo? Se mil participantes parece um número grande, na hora em que o total for dividido e fatiado para focalizar certos grupos, achados fundamentais às vezes residem em números *muito* pequenos.

Qual o tamanho do efeito? Se for um achado surpreendente, muitos epidemiologistas ignoram qualquer coisa que não tenha sido no mínimo duplicada em termos de risco/benefício de qualquer estudo observacional único. E se o risco inerente for pequeno, nem que seja duplicado, ele não é digno de preocupação.

Quão consistente é a ligação? Existe uma ligação convincente entre efeito e exposição?

Onde o estudo foi publicado? Ignore alegações feitas em conferências e aguarde a publicação num veículo científico respeitado. Mesmo então, lembre-se de que a publicação é uma condição necessária mas não suficiente para se impressionar. Veículos científicos de primeira linha podem publicar, e de fato publicam, bobagens.

No topo da hierarquia dos estudos observacionais encontram-se as imensas pesquisas prospectivas de coorte multicêntricas, capazes de controlar muitos fatores potencialmente indutores de erros, resultando em evidência convincente de fatores de risco plausíveis. Um exemplo clássico é o Million Women Study, montado em meados da década de 1990 por pesquisadores da Universidade de Oxford. Focalizado em mulheres com idade mínima de cinquenta anos, ele buscava vínculos entre sua saúde e uma miríade de fatores, desde uso de contraceptivos até dietas e tabaco. Em meados dos anos 2000, o estudo havia descoberto evidência de uma ligação entre risco de câncer de mama e uso de certos tipos de terapia de reposição hormonal (TRH). A ligação era ao mesmo

tempo forte e plausível, e o simples tamanho da coorte permitiu aos pesquisadores compensar muitos dos potenciais vieses sem minar a credibilidade de suas descobertas.

É inteiramente possível que durante as décadas por vir, estudos observacionais como o Million Women Study salvem milhões de vidas. Eles podem não ser tão confiáveis quanto o padrão-ouro do estudo clínico randomizado cego, mas as investigações prospectivas de coorte muito grandes e bem-administradas são boas o bastante. Mas, da próxima vez que você ler sobre algum risco implausível para a saúde, baseado num pequeno estudo de caso-controle, relaxe, respire fundo – e espere até ele ser derrubado.

Conclusão

Estudos observacionais nunca podem ser tão confiáveis quanto o padrão-ouro do estudo controlado randomizado duplo-cego. Mas frequentemente são a única maneira de lançar luz sobre questões críticas. E se forem abrangentes, bem administrados e seus resultados não forem forçados demais, também são dignos de confiança.

^e A expressão é *pecking order*, comumente empregada tanto em epidemiologia quanto em economia; é comum o conceito de “teoria da *pecking order*”. (N.T.)

12. Como uma “boi-bagem” deflagrou uma revolução

NINGUÉM SABE EXATAMENTE como foi construída a grande pirâmide de Gizé, mas pode apostar que demorou mais tempo e estourou o orçamento. Mais de 4 500 anos depois, essa é uma coisa que parece não ter mudado. Desde fazer o upgrade de um sistema de computadores até construir uma estação internacional no espaço, nenhum projeto é tão perfeito que sua realização se cumpra sem atrasos e gastos imprevistos.

Isso é estranho, considerando o esforço investido em métodos de gerenciar projetos planejados especificamente para impedir esses desastres. Com nomes impressionantes como Agile e PRINCE2, e jargão bizarro (*scrum of scrums*, SoS^f e *backlog grooming*^g), realmente soam admiráveis. Todavia, não está claro que funcionem, seja o que for que digam seus defensores.¹ Felizmente, a pesquisa hoje fornece alguma evidência bem convincente da eficácia de outro meio de prever o imprevisível. De modo irônico, ele tem origem numa questão que diz respeito a algo que pode realmente se chamar “boi-bagem”.^h

O boi em questão era um gigantesco macho, principal atração da Exposição de Animais de Gordura e Avícolas do Oeste da Inglaterra em Plymouth, Devon, em 1906. Os frequentadores eram convidados a usar sua habilidade de julgamento e estimar o peso do animal depois de abatido. Para tornar o desafio mais difícil, os organizadores pediam não o peso corporal vivo, mas o chamado “peso limpo” – ou seja, a massa da carcaça, menos cabeça, pés, órgãos e couro. Aproximadamente oitocentas pessoas pagaram o equivalente, na época, a mais ou menos £5 para participar, e, quando as estimativas foram examinadas, uma delas adivinhara corretamente o peso, mais ou menos 550 quilogramas. No entanto, outra pessoa se deu ainda melhor: o brilhante polímata Francis Galton. Ele decidiu descobrir simplesmente como tinha sido o desempenho das pessoas adivinhando o peso do boi, e obteve todos os cartões de palpites da competição.

Ao analisá-los, fez uma descoberta extraordinária. Ainda que a amplitude dos palpites fosse previsivelmente ampla, a mediana (isto é, o peso para o qual havia uma quantidade de palpites abaixo igual à quantidade acima) era 555 quilogramas – dentro da margem de 1% do peso real.

Como a adivinhação de todos aqueles indivíduos acabou produzindo um valor central tão próximo da verdade? O acaso era claramente uma possibilidade, mas, ao relatar sua descoberta na revista *Nature*, Galton sugeriu uma explicação mais intrigante. Julgou que a competição havia deflagrado um agrupamento de opiniões peritas. Segundo ele, a imposição de uma taxa de participação afastara muitos dos desocupados e desconhecedores, reduzindo aquilo que se poderia denominar “viés de ignorância”. Ao mesmo tempo, a perspectiva de ganhar incentivou os participantes habilitados a dar o melhor de si – aumentando ainda mais a possibilidade de exatidão. A combinação dos palpites individuais deu, portanto, uma estimativa coletiva baseada no conhecimento e na prática daqueles dispostos “a colocar seu dinheiro numa área de conforto”. E – no caso do peso do boi, pelo menos – o resultado foi impressionantemente exato.

Agora conhecido como “efeito da sabedoria das multidões”, essa interpretação continua controversa desde então – no mínimo porque parece violar regras básicas acerca de extrair conclusões a partir de informação limitada. Todavia, os céticos foram obrigados a encarar a crescente evidência de sua eficácia, tais como o êxito dos chamados mercados preditivos, que têm poderes capazes de assombrar o próprio Galton. No fim dos anos 1980, estudiosos da Universidade de Iowa estabeleceram o Iowa Electronic Market (IEM), no qual os entendedores podiam comprar e vender “ações” do resultado das eleições americanas. Os preços das ações refletiam as chances e a margem de vitória de cada candidato. Assim, por exemplo, se o preço da ação implicasse 80% de chance de um candidato vencer, mas alguém achasse que a chance real era de 85%, as ações pareciam uma boa aposta, e valia a pena comprá-las. Aqueles especialmente confiantes na sua crença estariam dispostos a adquirir montes dessas ações, fazendo assim com que o preço subisse – e daí a probabilidade implícita de vitória. Como os participantes estavam concentrados

em ganhar dinheiro, seu conhecimento especializado acabou se represando, revelando a sabedoria coletiva da “multidão” de entendedores.

No correr das décadas, essa multidão se mostrou espantosamente sábia. Uma análise de 2014, feita por dois pesquisadores da Universidade de Iowa, evidenciou que o IEM bateu os resultados das pesquisas de intenção de voto convencionais cerca de três quartos das vezes, com um erro de previsão para a ação do candidato nas eleições presidenciais americanas de apenas 1%. Desde então, o êxito do IEM tem se reproduzido em outros mercados preditivos. Fanáticos por cinema podem usar seu conhecimento para comerciar ações do sucesso de atores, novos lançamentos e ganhadores do Oscar no Hollywood Stock Exchange (HSX). Apesar de não oferecer nenhum incentivo maior que falsas fortunas de dólares e elogios, as predições da HSX têm se mostrado tão confiáveis que se montou um terminal para alimentar os palpites dos executivos da Cidade da Fantasia. Num celebrado exemplo, a sabedoria coletiva do HSX identificou o potencial de sucesso de um filme de horror com orçamento de US\$ 25 mil, que a gerência do estúdio ignorara. Chamava-se *A bruxa de Blair* e faturou quase US\$ 250 milhões de bilheteria.

A sabedoria das multidões também pode ser observada nas chamadas bolsas de apostas como a Betfair. Elas associam especialistas com opiniões opostas, e os ganhos de uma pessoa provêm das apostas perdidas por outras, sendo que a bolsa tira uma pequena fatia do lucro por ter organizado o bolo de apostas. Os especialistas são atraídos pelo fato de que geralmente conseguem vantagens maiores do que teriam de casas de aposta convencionais, cujos custos de administração mais elevados refletem-se em vantagens menos generosas. Mais uma vez, a pesquisa mostrou que a sabedoria das multidões refletida nas vantagens finais da bolsa é impressionantemente confiável: resultados cujas chances de ocorrer são consideradas pela multidão de, digamos, aproximadamente meio a meio realmente ocorrem cerca de 50% das vezes.

Como veremos em capítulos posteriores, essa acurácia aumentada das chances na verdade dificulta o sucesso como apostador. Mas mostra como a sabedoria das multidões pode produzir conclusões confiáveis mesmo em

situações complexas, envolvendo muitos fatores em interação. Isso não passou despercebido pelos encarregados do secular desafio de manter os projetos pontuais e dentro do orçamento.

No final da década de 1990, uma equipe da empresa multinacional de tecnologia Siemens resolveu descobrir se a sabedoria das multidões poderia se sair melhor que a administração convencional para manter um projeto de software na linha. Trabalhando com Gerhard Ortner, da Universidade de Tecnologia, em Viena, eles montaram um mercado preditivo possibilitando àqueles que trabalhavam no projeto comprar e vender “ações” cujo preço refletia as chances de o plano cumprir determinado prazo final. A equipe da Siemens estabeleceu dois mercados: um planejado para chamar a atenção para o risco de atraso, o outro para captar a percepção de sua provável duração. A esperança era de que os empregados fornecessem suas conclusões de forma rápida e anônima, via mercado, a fim de embolsar o lucro – dando assim um alarme precoce dos problemas. E foi exatamente isso que aconteceu. O mercado revelava no “preço” o impacto de mudanças no projeto muito antes de seu anúncio por parte da administração sênior, uma vez que os empregados corriam para se beneficiar de seus insights pessoais, e compravam ou vendiam ações. Em apenas um mês de negócios – e mais de três meses antes do prazo final em si –, os mercados prediziam que o prazo não seria cumprido, com um atraso estimado de duas a três semanas. Com um mês ainda faltando, os mercados foram inundados por um dilúvio de ordens de “vender”, sinal claro de que a confiança de cumprimento do prazo havia desabado. O projeto de software estourou o prazo, atrasando duas semanas. Entrementes, as ferramentas padronizadas de administração do projeto continuavam a insistir em que tudo correria bem até o prazo final.

Muitas empresas desde então fizeram experiência com os métodos da “sabedoria das multidões”. A Hewlett-Packard descobriu que mercados preditivos forneciam conjecturas mais confiáveis de vendas de impressoras que seus métodos convencionais de estimativa. A Google descobriu que eles ajudavam a prever a demanda futura de produtos tais como o Gmail, e possíveis ameaças à sua participação no mercado. Uma análise da performance

de seus mercados preditivos internos descobriu uma correlação impressionante entre as chances previstas de eventos segundo os mercados e a frequência com que os eventos de fato ocorriam. Ford, Procter & Gamble, Lockheed Martin, Intel, General Electric – a lista de corporações que têm usado mercados preditivos é bastante longa.

Então, se os mercados preditivos são tão maravilhosos, por que todo mundo não os usa o tempo todo? As razões são uma intrigante combinação de elementos racionais e irracionais. Sir Robert Worcester, fundador da empresa de pesquisa de mercado Mori, sem dúvida falava por muita gente ao caracterizar, em 2001, os mercados preditivos como “pseudopesquisas”. Sua preocupação estava focalizada na aparente violação das regras básicas da teoria de amostragem. Primeiro, mercados preditivos são qualquer coisa, menos amostras aleatórias; de fato, são planejados especificamente para terem o viés de incluir apenas aqueles confiantes para arriscar seu dinheiro ou reputação. Segundo, os mercados preditivos continuam bastante confiáveis mesmo quando envolvem apenas algumas dezenas de “negociantes” – um tamanho de amostra que a teoria-padrão condenaria como perigosamente pequena em muitas circunstâncias.

O mistério de como os mercados preditivos podem se safar com uma desconsideração tão flagrante pelas regras tem provocado muita controvérsia e pesquisa – e algumas pistas agora começam a surgir. Uma delas vem da experiência das empresas de pesquisa, que sabem muito bem que uma teoria que funciona bem com bolas coloridas nem sempre é digna de confiança quando se lida com pessoas vivas, reais. Com o correr dos anos, essas empresas viram sua metodologia aparentemente rigorosa ser jogada no lixo por pessoas que lhes dizem uma coisa e depois fazem outra. Elas tentaram vários artifícios para corrigir os efeitos dessas dissimulações, mas sem nenhum proveito óbvio.² Isso levou alguns pesquisadores a se perguntar se o efeito da sabedoria das multidões se beneficia do seu foco nas *características* dos indivíduos que formam a multidão.

Essa é uma ideia radical, semelhante a sugerir que se obtêm boas estimativas do que há dentro de um vaso de bolas coloridas se tiverem certa combinação de cores. E também tem implicações para a melhor maneira de se chegar às decisões coletivas. Mas é verdade? Pesquisas em campos tão diversificados quanto psicologia, administração, ecologia e ciência da computação têm demonstrado que, quando se trata de resolver problemas, decerto existe algo como ter quantidade demais de uma coisa boa. A questão não são os choques de personalidade nem egos demais; é simplesmente que um nível de qualificação elevado com frequência tem seu custo em termos de estreiteza. Em 2004, Lu Hong e Scott Page, da Universidade de Michigan, provaram matematicamente que um grupo de pessoas moderadamente qualificadas, com percepções diversificadas, em geral resolverão problemas com maior eficácia que uma equipe formada apenas por pessoas das mais altas qualificações.³

Isso tem ressonância óbvia no efeito da sabedoria das multidões. A conexão foi reforçada por uma equipe liderada pelo teórico de decisões Clinton Davis-Stober, da Universidade do Missouri.⁴ Eles começaram por captar matematicamente o conceito de sabedoria de multidão, e então examinaram o que pode miná-lo. Como Hong e Page, descobriram que a confiabilidade dos mercados preditivos depende das características daqueles que dele participam. Naturalmente, a habilitação desempenha seu papel, porém, mais uma vez, é a diversidade que emerge como crucial. Uma vez incluídos alguns entendedores num mercado preditivo, a matemática mostra que a confiabilidade melhora ao *não* se recrutar mais do mesmo, mas trazendo franco-atiradores que pensem de maneira distinta e/ou tenham acesso a diferentes fontes de compreensão. De fato, na realidade, vale a pena dar uma examinada nos níveis de habilitação dos novos recrutados apenas para obter uma diversidade maior. Isso porque os pontos de vista dos entendedores tendem a ser correlacionados, então, convocar maior número deles pode transformar pequenos vieses em importantes erros coletivos. Os pontos de vista dos franco-atiradores, em contraste, são, por definição, muito menos correlacionados entre si e com os dos outros. Logo, ainda que os vieses possam ser maiores, estão menos propensos a pressionar a visão coletiva final.

O trabalho de Davis-Stober e seus colegas é parte de um esforço contínuo maior para dar à sabedoria das multidões uma base teórica sólida. Ele mostra que a sabedoria coletiva pode se beneficiar até dos insights de amadores, e se mantém robusta mesmo quando alguns tentam deliberadamente provocar desvios no resultado. No processo, a pesquisa tem confirmado o valor de incluir os insights daqueles que – usando um clichê corporativo – “pensam fora da caixa”. E também tem lançado luz sobre por que a sabedoria coletiva pode vir à tona mesmo em grupos tão pequenos que mal merecem o apelido de “multidão”. Segundo Iain Couzin, da Universidade de Princeton, e do estudante de pós-graduação Albert Kao, a explicação reside na correlação – dessa vez entre as fontes de percepção usadas pelos que fazem parte da multidão.⁵ Se as fontes forem amplamente disponíveis, criarão correlações em meio àqueles que estão fazendo o julgamento – o que está bem, contanto que as fontes sejam confiáveis. Mas, se não forem, o julgamento represado de uma grade multidão estará propenso a ser dominado por essas correlações, levando a uma confiabilidade fraca. Em contraste, o julgamento médio de um grupo pequeno é menos preciso, mais diversificado – e, portanto, mais bem protegido contra ser solapado por insights falhos. Há, no entanto, um limite óbvio para esses benefícios, ao qual se chega com uma multidão de uma só pessoa. Ironicamente, os julgamentos de indivíduos há muito são reverenciados; de fato, sua fonte com frequência acaba com aquele venerado título de “guru”. Isso não quer dizer que nunca se deve confiar no guru; novas pesquisas têm identificado métodos que permitem até àqueles entre nós sem aspirações ao status de guru fazer julgamentos melhores (ver Box a seguir).

COMO FAZER ESCOAR A SABEDORIA DE SUA “MULTIDÃO INTERIOR”

Embora predições baseadas em crenças coletivas possam ser impressionantemente confiáveis, na realidade não precisamos de uma multidão para tirar proveito de sua sabedoria. Podemos fazer isso sozinhos – se tivermos o cuidado de incluir alguma variedade do tipo multidão no nosso pensamento. Stefan Herzog e Ralph Hertwig, do Instituto Max Planck para o Desenvolvimento Humano, vieram com uma técnica para fazer isso: “o autoempurrão dialético” (*dialectical boot-strap*).⁶ Felizmente, a coisa é mais simples do que parece. Primeiro,

venha com um palpite inicial do que você esteja querendo prever, usando todo e qualquer insight que tenha, e anote. Agora imagine que lhe dizem que isso está errado – e pense onde você pode ter pisado na bola. Que premissas não seriam inexatas, qual seria o impacto de mudá-las? A estimativa resultante seria mais alta ou mais baixa? Agora faça outra estimativa, com base em sua nova visão do problema. Herzog e Hertwig descobriram que a média dos dois palpites em geral está mais perto da resposta verdadeira que cada um deles individualmente.

Muitas das questões acerca da sabedoria das multidões ainda estão sendo investigadas – por exemplo, o tamanho ideal da multidão para diferentes problemas de julgamento, o papel do tipo de personalidade e os benefícios de se oferecer um feedback aos participantes. Mas uma coisa é clara: os céticos não podem mais alegar que a evidência para a sabedoria das multidões é puramente circunstancial. Agora existe um substantivo corpo de evidências observacionais cada vez mais respaldadas por uma teoria rigorosa. Além do mais, a suposta falta de evidências e teorias provavelmente nunca foi o motivo real do ceticismo. Muita gente simplesmente tem uma desconfiança visceral daquilo que para eles é uma tomada de decisão por parte das turbas. É verdade que as regras que governam a sabedoria das multidões contradizem a teoria mais familiar, e mesmo o senso comum. Ao contrário de amostras de bolas coloridas dentro de urnas, as pequenas multidões não são necessariamente menos confiáveis que as grandes. A importância “óbvia” do conhecimento também adquire mais nuances com a adição de mais franco-atiradores propensos a produzir melhor sabedoria coletiva do que recrutar novas “autoridades no assunto”.

Será que estamos prestes a assistir a uma revolução nas previsões, em que tudo, desde projetos de construção até política externa, é guiado pela sabedoria das multidões? Talvez, mas provavelmente não vale a pena pedir a opinião do seu guru de plantão.

Conclusão

Defrontado com o desafio de adivinhar alguma coisa, seja cauteloso e não acredite nos argumentos confiantes de qualquer indivíduo – não importa o grau de conhecimento que ele tenha do assunto. Em vez disso, monte um mercado preditivo (talvez mediante algum serviço on-

line, como o cultivatelabs.com) e convida todo mundo que tenha alguma opinião para alimentar seus insights em troca de dinheiro ou elogios. A pesquisa sugere que a sabedoria coletiva daí resultante se provará muito mais confiável que qualquer suposto “guru”.

^f *Scrum* é o nome que se dá à formação inicial num jogo de rúgbi, com os jogadores com os braços travados na altura dos ombros, empurrando o time adversário para obter a posse de bola; o jargão *scrum of scrums* é empregado no método Agile e refere-se ao grupo de pessoas encarregado de escalar as diversas equipes em diferentes áreas do projeto. (N.T.)

^g *Backlog* é a palavra usada para se referir a reservas, recursos a serem empregados; *backlog grooming* é aperfeiçoar, refinar esses recursos. (N.T.)

^h Adaptação para um trocadilho de difícil tradução. O termo empregado pelo autor é *a lot of bull* (“um monte de boi”), variação menos chula da consagrada expressão *a lot of bullshit*, com *bullshit* significando literalmente “merda de boi”; mas a expressão é utilizada para “conversa fiada”, “bobajada”, “baboseira” etc. Como se verá a seguir, a palavra *bull*, “boi”, aqui se justifica. (N.T.)

13. Como vencer os cassinos no jogo deles

NUMA NOITE DE SEXTA-FEIRA, agosto de 2014, Walter e Linda Misco, de New Hampshire, entraram no cassino MGM Grand, em Las Vegas, e rumaram diretamente para as brilhantes e reluzentes máquinas, verdadeiros ímãs dos perdedores, conhecidas como caça-níqueis. Desde a sua invenção, há mais de um século, esses “bandidos de um braço só” trocaram suas alavancas homônimas por botões e eletrônica, mas não perderam nada da capacidade de tirar dinheiro das pessoas. Isso não perturbou os Misco; na verdade, eles queriam encontrar a máquina mais conhecida da Cidade do Pecado: a Lion’s Share, ou Parte do Leão. Um dos caça-níqueis originais do MGM Grand tinha adquirido um infame renome mundial por jamais ter pagado uma só bolada desde que fora instalado, em 1993. Havia, porém, o outro lado dessa celebridade: sendo uma máquina do tipo “progressivo”, a mesquinhez da Lion’s Share significava que a bolada em oferta havia crescido a ponto de um eventual ganhador se tornar instantaneamente milionário. E, no correr dos anos, a máquina atraiu jogadores do mundo inteiro, que alegremente faziam fila para tentar a sorte.

Quando chegou a vez dos Misco, eles introduziram US\$ 100 para alimentar suas apostas, mais com esperança do que com verdadeira expectativa. Contudo, apenas cinco minutos depois de terem começado a sessão, três cabeças verdes de leão MGM apareceram em linha. As luzes piscaram, a máquina soltou um som estridente, e então ocorreu aos Misco que haviam feito o que ninguém conseguira antes: eles tinham ganhado a Parte do Leão: todos os US\$ 2,4 milhões.

Para muita gente, essa é uma das histórias mais reconfortantes, em que dona Sorte finalmente fez a coisa certa. Decerto foi assim que a mídia viu o episódio, e os Misco retribuíram revelando que tinham planejado usar o dinheiro para pagar a universidade dos netos e comprar um carro esporte. No entanto, para

outros, o que aconteceu com os Misco simplesmente ressalta tudo que há de errado nos cassinos e seus cínicos estratagemas para continuar a atrair os otários para dentro de suas portas.

Todo mundo tem uma opinião sobre cassinos. Alguns ficam fascinados com a imagem atraente, glamorosa, retratada em filmes como *Onze homens e um segredo* e *Casino Royale*. Outros são repelidos pela ideia de caloteiros dotados de máquinas que engolem economias da vida inteira. No entanto, para quem realmente quer entender de probabilidade, a visita a um cassino é obrigação. Eles são templos de astúcia probabilística. Com rendimentos acima de US\$ 150 bilhões por ano, os cassinos do mundo fornecem prova convincente acerca dos benefícios de se ter um ramo da matemática como núcleo de um modelo de negócios – especialmente um ramo de que a maioria das pessoas julga entender, mas não entende. Sua imagem pode estar maculada por associações com pessoas mais ansiosas para usar os punhos que o cérebro, porém, os cassinos devem seu sucesso ao uso inteligente do mais malcompreendido dos teoremas da probabilidade, a lei das médias. A maioria dos jogos bem conhecidos que eles oferecem, incluindo roleta, dados e caça-níqueis, tem resultados cujas probabilidades podem ser calculadas com precisão a partir de princípios básicos. Armados com esses princípios, os cassinos criaram um modelo pautado em prêmios que parecem razoáveis, mas não são. Eles são tudo, menos o que deveriam ser para um jogo genuinamente justo – mas, de forma esperta, a maioria deles não é injusta ao extremo. Trata-se de uma combinação que faz o notável truque de garantir que montes de jogadores se deem bem enquanto “a casa” ainda tem uma margem de lucro sólida como uma rocha.

Tomemos o jogo típico do cassino, a roleta, com sua famosa roda de 36 casas de números alternando vermelho e preto. Como há dezoito casas de cada cor, parece óbvio que a probabilidade de a bola cair numa casa vermelha ou preta é de 50:50; sem dúvida é isso que os cassinos querem que você pense, pois pagam valor igual à aposta para qualquer um que crave no vermelho ou no preto. Mas dê outra olhada na roleta: enfiada discretamente entre as casas vermelha e preta há outra, numerada com um “0” e colorida de verde; nos Estados Unidos em

geral há uma segunda casa verde, numerada “00”. Aquilo parece pouco importante, podem-se facilmente passar dezenas de giradas sem que a bola caia na casa verde. Mas uma soma rápida revela algo estranho. Suponha que você esteja num cassino de Las Vegas e aposte no vermelho. As chances de sair essa casa são dadas pelo número de casas vermelhas – dezoito – dividido pelo número total de casas em que a bola poderia cair, que é 38, pois temos de incluir as duas casas verdes como possibilidades. Logo, as chances de ganhar valor igual ao da aposta não são de 18 dividido por 36, mas 18 dividido por 38 – que é 47,37%, e não 50%.

Isso parece injusto – e é. Aquelas casas verdes fazem o jogo pender a favor do cassino. Mas aí está a coisa: o desvio é tão discreto – menos de 3% – que é facilmente engolido pelas flutuações aleatórias a curto prazo... como, por exemplo, o tempo gasto à mesa pela maioria dos jogadores. No decorrer de poucas horas, alguns poderão ganhar muito, outros irão amaldiçoar a sorte – mas ninguém será capaz de detectar o pequeno viés em favor da casa. De fato, a lei das médias mostra que ele se manifestaria de modo convincente só após uma observação cuidadosa de pelo menos mil giradas da roleta. Quem joga durante tanto tempo? Os cassinos, eis a verdade, com suas dezenas de roletas, 24 horas por dia, 365 dias por ano. Logo, ao mesmo tempo que o jogador individual não se sente trapaceado, a lei das médias assegura que os esforços coletivos conferem ao cassino uma margem de lucro sólida, ou “margem da casa”, de $2/38$ ou 5,3% de todas as apostas vermelho/preto (ou 2,7%, nos cassinos europeus).

Então, é possível bater os cassinos? Com os anos, muita gente tentou a sorte com várias estratégias simples, só para descobrir que a boa estrela acaba. Qualquer um familiarizado com a lei das médias sabe que truques como apostar no fim de “sequências” de vermelho não dão certo: a bola não tem memória do que fez antes, então as probabilidades continuam as mesmas a cada girada. Os cassinos ficam felizes de sugerir os supostos benefícios de se jogar de acordo com um “método de apostas” como o Martingale – basicamente dobrar ou sair – ou outros mais exóticos, como o sistema Labouchère ou o método D’Alembert (o fato de o matemático homônimo do século XVIII ter fracassado em entender

lançamento de moedas já conta tudo que você precisa saber sobre ele). Todos eles alegam combater os caprichos da sorte aumentando o valor das apostas quando as condições estão “favoráveis”, e reduzindo-as quando não estão. Podem dar lucro por algum tempo, mas no final todos fracassam pelos mesmos motivos. Primeiro, os cassinos não permitem que os apostadores fiquem aumentando as apostas segundo alguma estratégia; todos eles impõem um “limite da casa” para administrar sua exposição ao risco. E há a lei das médias, assegurando que, se você continuar a jogar, sentirá cada vez mais os danos causados pela margem da casa, não importa quão pequena ela seja. Essa combinação impede que qualquer “plano de apostas” transforme um jogo injusto num fluxo de renda confiável.

Mesmo assim, há maneiras de ganhar dinheiro em cassinos que não envolvem trapagens. Elas se baseiam em furos na aparentemente irretocável lei das médias. Lembre-se de que a lei afirma que a probabilidade de um evento resultante de um processo aleatório pode ser estimada de forma ainda mais precisa dividindo-se o número de vezes que ocorre pelo número sempre crescente de oportunidades de ocorrer. Assim, por exemplo, na roleta, a proporção de vezes que a bola cai no vermelho ficará cada vez mais próxima do valor teórico de 47,37% à medida que a quantidade de giradas aumenta.

No entanto, sorrateiramente à espreita no resultado matemático irretocável, há diversos senões. O mais óbvio é a premissa de que o processo propulsor do jogo seja realmente aleatório. Como vimos no caso do lançamento da moeda (Capítulo 1), o que parece aleatório e imprevisível pode na realidade ser extremamente complicado e, pelo menos em termos amplos, previsível. No caso da roleta, a bola saltando e rebatida está, em última análise, sujeita às leis da física, e como tal seu movimento não pode ser genuinamente aleatório, o que por definição significa não obedecer a nenhuma regra.

Esse furo na lei das médias já respaldou muitas tentativas bem-sucedidas de tirar dinheiro dos cassinos. Como engenheiro trabalhando na indústria de algodão vitoriana, Joseph Jagger sabia que os dispositivos mecânicos nem sempre funcionam exatamente como se pretende. Isso o levou a indagar se

haveria falhas na operação das roletas que se pudessem explorar. Em 1873, ele mandou uma equipe para Monte Carlo a fim de monitorar sub-repticiamente a performance das roletas no Cassino des Beaux Arts. Efetivamente, eles descobriram que as bolas apresentavam maior propensão a cair em alguns setores da roda que em outros. O viés era pequeno demais para ser identificado pela gerência do cassino, mas – crucialmente – grande demais para superar a estreitíssima margem da casa em algumas apostas. Isso tornava as chances ligeiramente injustas em jogadas lucrativas para certas apostas em determinados números. Armado desse conhecimento, Jagger foi a Monte Carlo e durante alguns dias, em julho de 1875, ganhou o equivalente a £3 milhões em dinheiro atual – o que fez dele o melhor candidato na vida real ao título de “Homem que quebrou a banca em Monte Carlo”.

Os cassinos perceberam desde então a importância de checar com regularidade todos os seus equipamentos em busca de defeitos, desgaste, partes quebradas e mau funcionamento. Mas isso não fecha de todo o furo, pois até uma roleta nova em folha, perfeitamente ajustada, está sujeita às leis da física, que oferecem pelo menos alguma previsibilidade. Em 1961 os matemáticos Claude Shannon e Ed Thorp – indiscutivelmente as duas melhores cabeças a enfrentar os cassinos – construíram um computador capaz de transformar observações de como e onde uma bola fosse jogada na roleta em previsões dos quatro ou cinco números em que ela cairia. Isso transformou a diminuta margem de lucro do cassino em robustos 40% a favor de Shannon e Thorp. Problemas técnicos impediram a dupla de levar o aparelho ao cassino, mas a ideia foi revivida no fim dos anos 1970 por um time de estudantes de física da Universidade de Santa Cruz. Eles encaixaram numa bota de caubói um microprocessador, foram para Las Vegas e alegaram ter ganhado um lucro respeitável.

A estratégia de usar as leis da física para aproveitar os furos na lei das médias hoje se combina com uma tecnologia cada vez mais sofisticada. Em março de 2004, um trio da Europa Oriental se apossou de £1,3 milhão no cassino Ritz, em Londres, usando um laser escondido num telefone celular falso a fim de

compilar os dados necessários para prever onde a bola cairia. Depois de analisar gravações em vídeo, o cassino chamou a polícia, mas o trio foi liberado sem acusações, com a permissão de conservar o que tinham recebido.

Outro furo, mais sutil, na lei das médias permite ganhar dinheiro jogando um dos jogos de cartas mais populares nos cassinos: o blackjack ou vinte e um. Explicando de forma simples: jogadores e banca recebem cartas, e os jogadores apostam que conseguem chegar a um valor mais perto de 21 – ou exatamente 21 (blackjack) – que a banca, na soma do número das cartas. As regras variam, mas em geral elas fazem com que o jogo seja injusto, embora a margem da casa seja estreitíssima, de menos de 1%. Contudo, há um furo enterrado no cálculo dessa margem que os jogadores habilidosos conseguem explorar. As cartas são distribuídas de diversos baralhos misturados – na maior parte das vezes, meia dúzia – e depois descartadas, e não devolvidas ao todo (processo que os matemáticos chamam de “amostragem sem reposição”). Desse modo, enquanto valores específicos de cartas podem aparecer ao acaso, não há um suprimento infinito delas; se estão, digamos, quatro baralhos em jogo, uma vez tendo aparecido dezesseis ases, você não verá mais nenhum ás até o conjunto todo voltar a ser embaralhado. Isso quer dizer que as chances de ter mãos vencedoras no blackjack – ao contrário de outros jogos de cassino, como a roleta – não são fixas, mas mudam à medida que o jogo avança. Isso torna mais suaves as garras da lei das médias, permitindo que as chances de ganhar se voltem significativamente em favor dos jogadores. Melhor ainda, também solapa a regra de que não existe meio de transformar um jogo injusto em jogo lucrativo simplesmente apostando de um jeito específico. No blackjack, você se contém enquanto as chances de ganhar estão contra você, e entra com tudo quando elas ficam a seu favor.

Identificar quando isso ocorre envolve a técnica conhecida como “contagem de cartas”. Concebida pelo matemático Ed Thorp, que divulgou a técnica no seu best-seller de 1962, *Beat the Dealer* (até hoje reeditado), a contagem de cartas foi inicialmente menosprezada pelos cassinos como mais um esquema tipo “fique rico depressa”. Mas a verdade foi que se deram mal: acharam que

embaralhar as cartas era suficiente para garantir a margem da casa. Não levaram em conta o fato de que o ato de jogar revela a identidade das cartas que saem do baralho, e isso permite perceber o que tem probabilidade de acontecer em seguida.

Thorp concebeu um sistema para acompanhar quais cartas já tinham saído e, segundo isso, ajustar as apostas. Como o impacto da contagem de cartas é relativamente pequeno, ele exige um robusto saldo bancário e concentração contínua para transformá-la em lucro decente. Mesmo assim, a publicação do livro de Thorp fez com que os cassinos perdessem quantias substanciais para muita gente, desde estudantes universitários até aposentados que se deram ao trabalho de dominar a contagem. Então os cassinos revidaram. Começaram a aumentar a quantidade de baralhos usados para no mínimo seis – ampliando a exigência mental da contagem. Depois introduziram os embaralhadores automáticos, que embaralham as cartas no meio de uma sessão, jogando no lixo as contagens em andamento. A rapidez desses embaralhadores também fez crescer o número de apostas por hora, dando assim à casa mais tempo para fazer sua magia. Muitos cassinos simplesmente mudaram a proporção do pagamento-padrão das apostas no blackjack, cancelando dessa forma a minúscula vantagem de contar as cartas.

Apesar de tudo isso, ainda há por aí contadores de cartas em profusão, e para eles os cassinos reservam a contramedida definitiva: a “chamada discreta”. Embora não seja ilegal, a contagem de cartas, ainda que apenas suposta, é considerada inaceitável pela maioria dos cassinos – e eles não se importam com quem saiba disso. Conta-se que em 2014 o astro de Hollywood (e bem-sucedido jogador de cartas) Ben Affleck teria recebido uma “chamada discreta” da gerência do Hard Rock Cassino, de que ele era bem-vindo para jogar qualquer outro jogo – o jeito Vegas de dizer: “Achamos que você está contando cartas e vai ter de parar.”

Indiscutivelmente, a estratégia única mais efetiva para ganhar uma fortuna no cassino é deixar claro que você já é riquíssimo – o que o torna uma “baleia”, na linguagem de Las Vegas. Cassinos adoram as baleias, pois elas gastam muito

e perdem ainda mais, porém, podem cobrir suas dívidas. Por conseguinte, os cassinos atendem alegremente a qualquer capricho que a baleia possa ter. Era isso que o magnata do jogo e especialista em blackjack Donald Johnson esperava que acontecesse quando deu um golpe em vários cassinos de Atlantic City, em 2011, com resultados espetaculares. Deixando claro que jogaria mãos de US\$ 25 mil, ele conseguiu negociar uma porção de pequenas alterações nas regras padronizadas do blackjack, todas elas reduzindo a vantagem da casa. Aí usou as duas estratégias normalmente associadas aos velhos cassinos dirigidos pela Máfia: diversionismo e intimidação. Johnson aparecia a cada jogo com um destacamento de mulheres trajadas de roupas provocantes. A presença delas, além da enervante quantia apostada por Johnson, fazia com que o responsável pela banca perdesse a concentração e cometesse erros. Isso levou os cassinos a liberar apostas para Johnson, que afinal virou a vantagem a seu favor.

Ao longo de vários meses, Johnson levou sua estratégia a diversos cassinos de Atlantic City e os depenou em cerca de US\$ 15 milhões. Seu golpe foi tema de manchetes, gerentes foram demitidos – até vir a “chamada discreta” de muitos cassinos, dizendo-lhe que ele não era mais bem-vindo.

Então, a velhíssima alegação é verdade: há meios de vencer os cassinos no próprio jogo deles. A má notícia é que isso envolve precisamente isto: ter níveis proporcionais de habilidade, determinação e dinheiro. Mas a maioria das pessoas que frequentam cassinos não planeja fazer disso uma carreira; elas vão pelo divertimento e a sedução de talvez ganhar um pouquinho de dinheiro. A boa notícia é que – como veremos no próximo capítulo – as leis da probabilidade levam a algumas dicas de primeira para maximizar as chances de ter as duas coisas.

Conclusão

Cassinos são fábricas que usam as leis da probabilidade para produzir lucro. Furos nessas leis realmente possibilitam desviar um pouco desse lucro para você, mas são pequenos e penetrá-los requer habilidade, determinação e um bocado de dinheiro.

14. Onde os espertinhos se dão mal

QUANDO OS MISCO SAÍRAM do MGM Grand em Las Vegas com US\$ 2,4 milhões, eles não pensaram em qualquer alternativa além de ter tido sorte. Simplesmente aconteceu de jogarem no caça-níqueis Lion's Share no dia em que a máquina pagou a primeira bolada em 21 anos. Durante esse tempo, ela rendera ao cassino mais de US\$ 10 milhões de lucro, num fluxo de renda respaldado pela lei das médias. Claro, o que quer que os Misco tenham recebido, com certeza não era a parte do leão da quantia que o caça-níqueis tirara das pessoas. Nenhum jogador sério nem chegaria perto das máquinas caça-níqueis, com sua enorme margem de 5 a 15% para a casa e competência de habilidade zero. Em vez disso, eles se concentram em jogos de baixa margem, como blackjack e bazará, planejando usar seu talento em estratégias como contagem de cartas para obter algum lucro.

Contudo, até os jogadores mais espertos podem cair na armadilha de pensar que estão ganhando por sua habilidade, enquanto é apenas a lei das médias que lhes concede um pouco de seu tempo. E se estão jogando um jogo com uma inevitável margem da casa, como o bazará, cedo ou tarde a lei acabará cobrando o tempo concedido. Saber quando largar o jogo é, portanto, uma habilidade-chave de qualquer jogador profissional. Mesmo assim, ela pode iludir mesmo os jogadores mais perspicazes.

Como magnata japonês do ramo imobiliário, Akio Kashiwagi era ao mesmo tempo esperto e rico. E também viciado em bazará, com um estilo implacável que lhe valeu o título de “O Guerreiro”. Não achava nada de mais participar de jogos com cacife de US\$ 100 mil que se estendiam por dias inteiros. Entre os gerentes de cassino, Kashiwagi era uma “baleia”: rico, confiante e disposto a apostar. Por isso, era “paparicado” em pródiga escala: comida, bebida, quartos VIP, até voos de ida e volta ao cassino eram agrados oferecidos pela gerência. O objetivo da casa era simples: mantê-lo ali tempo suficiente para ser destruído

pela lei das médias. Ao contrário do blackjack, o bazará tem uma vantagem da casa que não pode ser revertida por jogos ou apostas habilidosos. Em 1957, dois matemáticos haviam encontrado um modo ideal de jogar bazará, mas tudo só servia para adiar o inevitável; jogue o tempo suficiente, e a lei das médias pega você.

E assim foi que, em maio de 1990, Kashiwagi sentou-se a uma mesa de bazará arranjada especialmente para ele no recém-aberto cassino Trump Taj Mahal, em Atlantic City. As apostas eram de US\$ 200 mil a mão, e o jogo deveria continuar até ele ou o cassino ganhar US\$ 12 milhões. Kashiwagi correspondeu à sua reputação, jogando com habilidade e persistência, chegando a acumular US\$ 10 milhões. Mas então a estreitíssima margem começou a se voltar contra ele, e Kashiwagi cometeu o clássico erro de todos os jogadores inveterados: começou a querer compensar suas perdas. À medida que se passavam as horas, estas foram se acumulando. Finalmente, após setenta horas de jogo ao longo de seis dias, ele pegou US\$ 2 milhões em fichas e foi embora.

Mas então a estratégia do Taj começou a se desenrolar. O cassino apostou que Kashiwagi valia os US\$ 10 milhões que estava devendo. Mesmo assim, ainda havia um prejuízo de US\$ 6 milhões em janeiro de 1992, quando Kashiwagi foi encontrado morto em sua casa, perto do monte Fuji. Ele fora esfaqueado mais de cem vezes – alguns acreditam que por ordem da Yakuza, equivalente japonesa da Máfia. De forma bizarra, ele acabou adquirindo uma espécie de imortalidade numa cena do filme *Cassino*, de Martin Scorsese, de 1995. Alguns dos detalhes foram alterados – Kashiwagi em Atlantic City tornou-se “K.K. Ichikawa” em Las Vegas –, mas o desfecho e a moral eram iguais. Ele começa ganhando, porém fica muito ambicioso, joga bazará tempo demais – e sofre as consequências. As palavras do gerente ficcional do cassino, Sam Rothstein (baseado no chefe de cassino real Frank “Lefty” Rosenthal), deixam clara a estratégia usada pela casa para atrair uma baleia: “A regra cardeal é mantê-los jogando e fazer com que sempre voltem. Quanto mais tempo jogarem, mais vão perder. No final, nós ficamos com tudo.”

No entanto, os cassinos precisam de mais que baleias para ter sucesso, e até a margem da casa mais robusta não vale nada se não houver jogadores entrando pelas portas. Essa verdade básica propiciou uma reviravolta final para a história da baleia à mesa de bacará. Em 2014, cinco dos maiores cassinos de Atlantic City fecharam por falta de movimento; entre elas estava a nênese de Kashiwagi, o Taj Mahal.

A maioria das pessoas que vão aos cassinos não é formada por baleias, mas mesmo assim correm o risco de sucumbir à sedução de se aventurarem fora de suas profundezas. É vital saber como identificar sinais de perigo, como tirar o máximo das “iscas” apresentadas pelos cassinos e que iscas evitar de todo. Isso significa aplicar as leis da probabilidade. Ao mesmo tempo que a matemática por trás dessas leis é surpreendentemente complexa e ainda provoca controvérsias, ela é fácil de aplicar, e faz sentido intuitivamente.

A primeira lei explora o fato de que a aleatoriedade em geral leva tempo para se revelar. Lance uma moeda algumas vezes, e é perfeitamente possível tirar só caras ou só coroas, sugerindo que a moeda não se comporta de maneira aleatória. Entretanto, continue os lançamentos, e o fato de haver dois resultados possíveis vai se tornando cada vez mais claro. Isso é um sintoma daquilo que os matemáticos chamam de natureza “assintótica” da lei das médias – ou seja, o que ela afirma acerca das frequências relativas aplica-se estritamente a uma sequência infinitamente grande de eventos. Para qualquer sequência finita, toda uma gama de possibilidades é consistente com a aleatoriedade e pode ser radicalmente diferente da média de longo prazo para sequências curtas.

Quando aplicado aos jogos de cassino, isso significa que, durante sessões breves, podem-se obter afastamentos bastante significativos da margem da casa, ou margem de lucro – e se, para começo de conversa, essa margem da casa for bastante estreita, o resultado será uma explosão de lucratividade para os jogadores. A sessão mais breve de todas é obviamente uma jogada única. Embora isso não mude as chances a seu favor, minimiza o tempo que você fica exposto à lei das médias – e portanto o tempo durante o qual a margem da casa se faz sentir. Essa estratégia da jogada única foi adotada por Ashley Revell na

espetacular aposta vencedora de US\$ 135 mil descrita na Introdução. Ele foi esperto em jogar uma só vez – mas também foi sortudo.

Esse jogo arrojado não é para os fracos de coração, nem é muito bom para os ávidos por vivenciar a atmosfera de um cassino. Então, é necessário um meio-termo, e a melhor coisa é procurar jogos com a menor margem da casa e jogar tempo suficiente para ter uma boa chance de se sair bem, mas não tempo bastante para que a lei das médias comece a agir.

Para alcançar o primeiro objetivo, evite máquinas caça-níqueis ou jogos de loteria como keno, cujas atraentes boladas são financiadas exatamente pelas atraentes margens da casa. Em vez disso, focalize em apostas simples na roleta (como vermelho/preto), ou aprenda a jogar e explorar as apostas de baixa margem em jogos como blackjack e dados. Em seguida, resolva quanto tempo e dinheiro você tem para gastar no cassino, e jogue só até que um dos dois tenha se esgotado. Mas não passe seu tempo fazendo montes de pequenas apostas, pois isso reduz as chances de se dar bem. Por exemplo, suponha que você entre num cassino com £100 e decida tentar a sorte na roleta. Dependendo do movimento na mesa, são mais ou menos de trinta a quarenta jogadas por hora. Você tem mais chance de pelo menos sair em casa se passar quinze minutos fazendo apostas de £10 do que meia hora fazendo apostas de £5. Isso porque, no primeiro caso, você fará apenas dez apostas, no segundo, fará vinte – e, dividindo pela metade sua exposição à margem da casa, você aumenta as chances de lucrar £50 antes de passar do risco de 1 entre 3 para cerca de 50 a 50%. Se fizer umas dez apostas vermelho/preto, matematicamente terá chance maior do que 50 a 50% de pelo menos sair em casa, quase 1 chance em 3 de sair com algum lucro – e uma chance de 100% de dizer que jogou roleta e sabia o que estava fazendo.

Tampouco seja ambicioso demais em suas metas. Não resolva insistir até ter dobrado seu dinheiro. Os objetivos mais modestos têm maior chance de se atingir. Assim, por exemplo, enquanto você possui uma chance de 50 a 50% de transformar £100 em £150 antes de ser depenado jogando vermelho/preto com apostas de £10, essas chances são cortadas pela metade se o seu objetivo forem £200. E, claro, não caia em nenhuma conversa fiada sobre “aproveitar a sua

sorte” se tiver alcançado a meta em pouco tempo. Pegue o dinheiro e dê o fora – antes que a besta da aleatoriedade desperte e devore tudo.

Siga essas regras,¹ e você terá uma chance maior de voltar de sua visita a um templo da sagacidade probabilística com um sorriso no rosto.

Conclusão

Até jogadores profissionais podem confundir sorte com habilidade e passar tempo demais no abraço mortal da lei dos grandes números, tentando aumentar seu lucro ou recuperar as perdas. O truque para se divertir nos cassinos é aumentar a disciplina, reduzir a ambição e cortar as perdas.

15. A regra áurea das apostas

SE OS CASSINOS REPRESENTAM o lado glamoroso da jogatina, as casas de aposta dos becos são a sua antítese. De aparência enganosa, soturnas e levemente ameaçadoras, há muito têm sido notórios antros de caloteiros e desesperados. Contudo, também são testemunhas da popularidade de uma forma de jogo que faz parecer diminutos os jogos como roleta e blackjack. Trata-se das apostas em eventos esportivos: apostar no resultado de qualquer coisa, desde o cavalo que vai ganhar o Grande Prêmio Nacional até quantos panos amarelos serão jogados durante uma partida de futebol americano.ⁱ

As apostas nos esportes são um empreendimento global imenso, gerando rendas estimadas em cerca de US\$ 1 trilhão por ano. Na cidade de Hong Kong, só as corridas de cavalo produzem um giro de US\$ 10 bilhões. Apostas em eventos esportivos únicos, como o Superbowl, a final do futebol americano, têm alcançado níveis semelhantes.

Essas somas estarrecedoras depõem sobre o fato de que centenas de milhões de pessoas apreciam a “adrenalina” ocasional, respaldando nossas crenças em dinheiro sólido. Com o advento das apostas pela internet, nunca foi tão fácil jogar. Bet365, o maior site de apostas on-line da Grã-Bretanha, viu mais de £34 bilhões fluírem pelas suas apostas em 2015. Apesar de serem encaradas com o cenho franzido pela sociedade educada, nenhum volume de broncas foi capaz de impedir a crescente popularidade de se apostar. Todavia, aqueles que consideram direito inalienável jogar num cavalo no Jockey Club devem encarar o fato de que os apostadores mais regulares nos esportes perdem muito dinheiro – às vezes com consequências desastrosas. Embora seja tentador jogar a culpa nos agentes de apostas, o motivo real é claro como o dia: os apostadores mais regulares não entendem realmente o que fazem. Podem saber como ler um formulário e preencher cada tipo de aposta, mas não têm ideia de como distinguir uma aposta

decente para transformá-la em lucro. O que, obviamente, é a maneira de os agentes gerarem o seu próprio lucro.

Estima-se que cerca de 95% daqueles que apostam em eventos esportivos não conseguem obter um lucro consistente.¹ Então, o que os outros 5% sabem que todo o resto não sabe? Para grande surpresa, não é nada muito complicado; de fato, é de admirar que tão poucos estejam a par. Só aqueles que realmente tentaram entendem como um princípio tão simples pode arrasar sua sanidade.

Poucos dominaram a arte e a ciência de apostar quanto Patrick Veitch,² apostador inglês em corridas de cavalo. Ele se tornou multimilionário a partir de suas pesquisas, mais o título de Inimigo Número 1 das casas de aposta britânicas. Contudo, o histórico do seu sucesso deveria servir de advertência para qualquer um que sonhe em imitá-lo.

Veitch é, antes de tudo, extremamente inteligente. Já aos quinze anos conseguiu um lugar no Trinity College, em Cambridge – a *alma mater* de Isaac Newton – para estudar matemática. Impedido de iniciar seus estudos formais por causa da idade, passou o final dos anos 1980 afiando suas habilidades como apostador sério. Seu foco logo recaiu sobre as corridas de cavalo, atraído não só pela imagem ou a popularidade, como também pela sua *complexidade*. A chance de vitória de um cavalo numa corrida depende de um sem-número de fatores, desde as performances passadas e a qualidade dos concorrentes até o formato e a transitabilidade da pista naquele dia. Nada tendo a ver com o desafio intelectual, o adolescente Veitch já tinha identificado algo que sempre escapa da maioria dos apostadores: a complexidade lhe dava maior chance de discernir fatores não levados em consideração por todos os outros – incluindo as casas de aposta ao elaborar seus “diagramas” de chances para cada corrida. Esse foi um sinal precoce da estratégia de apostas que se tornou a base da fortuna de Veitch.

Uma vez no Trinity, Veitch rapidamente se distinguiu em matemática aplicada, embora não do tipo estudado pelos outros alunos. Enquanto estes permaneciam sentados durante as aulas de cálculo vetorial, Veitch ia a programas de corridas, fazendo rotineiramente apostas de £1 mil. Começou então a oferecer um serviço de dicas, tão bem-sucedido que, no começo do

último ano de faculdade, Veitch tinha £10 mil por mês fluindo pelas suas apostas. Concluiu que perdia tempo estudando matemática e largou Cambridge antes de se formar.

É possível que Veitch nunca tenha ido às aulas da graduação sobre probabilidade. Se tivesse ido, nunca teria encontrado as provas-padrão da lei das médias (ou lei fraca dos grandes números, como é inutilmente chamada pelos professores) nem aprendido sua implicação: a longo prazo, a probabilidade real de qualquer evento ao acaso é revelada com precisão cada vez maior pelo número de vezes que o evento ocorre, dividido pelo número de oportunidades que ele teria de ocorrer. Sem dúvida os alunos recebiam folhas de problemas para resolver com exercícios sobre a probabilidade de certos resultados em lançamentos de moedas ou de dados. No entanto, tudo isso teria pouco interesse ou utilidade para Veitch, porque focalizava o tipo errado de probabilidade.

A ideia de que existem diferentes tipos de probabilidade tem provocado há séculos amargos debates entre os estudiosos. Encontraremos algumas das infelizes consequências dessa controvérsia em capítulos posteriores. Ela já gerou um bocado de expressões (probabilidade “aleatória” *versus* “epistêmica”, frequentismo *versus* bayesianismo), além de divagações filosóficas e matemáticas. Mas a noção básica de diferentes formas de probabilidade é fácil de captar, mediante a diferença entre cassinos e casas de aposta. Os cassinos conhecem as chances de todos os vários resultados em jogos como roleta, dados e caça-níqueis. As probabilidades não precisam ser adivinhadas ou estimadas a partir de dados brutos. Podem ser encontradas a partir dos primeiros princípios. Há 38 casas nas quais uma bolinha de roleta em Vegas pode cair, então, a chance de cair numa delas é exatamente igual à chance de cair em qualquer outra. Essa é a probabilidade aleatória desse evento, e permite que os cassinos saibam que a lei das médias fará sua mágica a favor deles. As casas de apostas, em contraste, não têm as mesmas garantias, simplesmente porque não é possível calcular a probabilidade de, digamos, um cavalo ganhar a corrida a partir dos primeiros princípios. Ao contrário do giro da roleta, o resultado da corrida depende de uma mistura complexa de variáveis, desde o estado físico do cavalo, passando pelo

jóquei, até o estado da pista. Assim, as casas de apostas precisam se apoiar em seu próprio julgamento acerca das chances de um cavalo (sua “probabilidade epistêmica”, em jargão) e usá-lo para estabelecer seus critérios.

UM JEITO ÍMPAR DE FALAR

Apostadores em eventos esportivos estão nessa pelo dinheiro (pelo menos em teoria), então, tradicionalmente, descrevem as chances dos eventos não como probabilidades, mas como o lucro gerado por uma aposta vencedora. Assim, por exemplo, em vez de dizer que um cavalo tem 22% de chance de ganhar, dizem que ele paga “7 para 2”, querendo dizer que, para cada £2 apostadas, o ganho justo para uma vitória seria de £7. Para converter chances do tipo “X para Y” para uma probabilidade em termos de porcentagem, divida Y por X + Y e multiplique por 100. Para eventos de alta probabilidade, os apostadores falam de um evento de “3 para 1”, referindo-se a apenas £3 de ganho para cada £1 apostada. Para converter essas porcentagens, basta trocar o X e o Y da fórmula – assim, por exemplo, “3 para 1” torna-se 75%.

No entanto, o que as casas de apostas têm em comum com os cassinos é a determinação de lucrar oferecendo prêmios um pouco menos generosos do que deveriam. Para ver como isso funciona, suponha que os responsáveis por estabelecer as proporções numa casa de apostas acreditam que um cavalo tenha 40% de chance de ganhar (“6 para 4”, na linguagem dos apostadores, o que significa que uma vitória dê um ganho de £6 para cada £4 apostadas – ver Box abaixo). As chances reais anunciadas pela casa de apostas não serão de 6 para 4, porém, algo mais perto de “elas por elas”, implicando uma chance de 50% de ganhar. Como o nome deixa claro, elas por elas paga £4 para cada aposta de £4 – o que é muito menos generoso que uma chance de 6 para 4. Em outras palavras, o prêmio oferecido é injusto com o apostador, e a casa embolsa a diferença como lucro. Quem pensar que os prêmios pagos pela casa de apostas refletem acuradamente as chances de um evento ocorrer cai direitinho na armadilha. As chances divulgadas são o equivalente do truque do cassino, aparentando oferecer um prêmio justo, quando na verdade não faz nada disso – a diferença é que a

margem de lucro (às vezes chamada *overround*, ou “excedente”) é de 20% ou mais.

Isso pode parecer um modelo de negócios bastante lucrativo, mas é muito menos confiável que a margem dos cassinos, porque as chances baseiam-se em julgamento – e uma corrida de cavalos, ou na realidade qualquer evento esportivo, pode falhar em seguir o roteiro. As casas de apostas tentam se proteger contra isso oferecendo proporções de pagamentos injustas para cada resultado possível do evento – digamos, uma vitória em casa, ou na casa do adversário, ou empate para um jogo de futebol. Ao mesmo tempo que precisam estabelecer um equilíbrio entre o que os competidores oferecem e os apostadores aceitam, seu objetivo é dar a si mesma uma “aposta equilibrada”, com boa chance de produzir uma margem de lucro decente, seja qual for o resultado.

Tomemos o caso real das proporções oferecidas por uma casa de apostas num jogo qualificatório para a Euro 2016, com as chances convertidas em probabilidades para cada resultado:

RESULTADO	VITÓRIA DA INGLATERRA	VITÓRIA DA ESLOVÊNIA	EMPATE
Proporção nas apostas	4 para 11	10 para 1	4 para 1
Probabilidade correspondente	73%	9%	20%

Tudo isso parece fazer sentido. Há somente três resultados possíveis para o jogo: vitória de um dos dois times ou empate – e todos receberam suas proporções nas apostas, sendo que a Inglaterra tem mais chance de vitória que a Eslovênia, embora o empate seja possível. Mas observe melhor, e torna-se evidente o efeito da determinação da casa de apostas de ter seu lucro. Como um dos três resultados precisa necessariamente acontecer, as chances individuais deveriam somar 100%. Entretanto, o total na aposta do exemplo perfaz 102%. Esse é o sinal denunciante de que a proporção de prêmios oferecidos não representa a crença real da casa acerca das chances de cada resultado, pois estas teriam de somar 100%. Em outras palavras, as chances reais para pelo menos um

dos resultados, na opinião da casa, são mais baixas que as apresentadas – e a diferença de 2% é embolsada como lucro.

Para ser justo, as casas de apostas precisam de considerável habilidade para estabelecer até essas chances injustas, pois apenas se estiverem baseadas numa estimativa acurada das chances reais poderão se tornar fonte de lucro. Se os responsáveis por estabelecer essas proporções errarem em suas estimativas das chances reais, acabarão inadvertidamente oferecendo prêmios muitíssimo generosos. É aí que entra gente como Veitch e outros apostadores bem-sucedidos. Eles usam sua própria habilidade para estimar as chances reais de cada resultado, e então as comparam com as chances oferecidas pelas casas de apostas. Seu objetivo é descobrir as chamadas “apostas de valor” – ocorrências em que as casas de apostas deixaram de ver algo crucial na sua análise, e portanto oferecem prêmios generosos demais.

O que eles fazem requer habilidade e determinação enormes, mas em essência pode ser resumido numa fórmula simples, que poderia ser chamada de regra áurea das apostas (ver Box a seguir).

A REGRA ÁUREA DAS APOSTAS

Ganhar regularmente dinheiro com apostas exige um método comprovado para identificar “apostas de valor”. Estas exigem que as verdadeiras chances de um evento ocorrer sejam significativamente *mais altas* do que sugerem as proporções das casas de apostas. Identificar apostas de valor demanda, portanto, a descoberta dos fatores que afetam resultados que mesmo as casas de apostas não levaram totalmente em consideração ao estimar as chances reais do evento. Sem um método comprovado de encontrar e explorar esses fatores, as apostas poderão eventualmente produzir perdas substanciais.

A regra áurea simplesmente cristaliza o fato de que as proporções oferecidas pelas casas de apostas não podem ser tomadas pelo seu valor nominal. Elas foram deliberadamente “maquiadas” para pagar *menos* do que realmente deveriam, à luz da estimativa da casa acerca da verdadeira probabilidade de ocorrência do evento, em geral considerada mais baixa que as chances reais.

Assim, qualquer um que se baseie nas proporções das casas de apostas para avaliar as chances de ganhar acabará amargando uma pesada perda.

Claro que se você estiver apenas fazendo apostas ocasionais em eventos grandes para se divertir um pouco, nada disso tem muita importância. A diferença entre as chances reais e as divulgadas de hábito é pequena o bastante para ser tomada ao menos como um guia aproximado do ranking relativo de vários resultados. Favoritos que pagam pouco realmente tendem a ganhar com mais frequência que os penetras mal ranqueados. Mas o perigo surge se você resolve fazer tantas apostas só por diversão que a diferença comece a se revelar, à medida que o efeito de longo prazo da lei das médias se manifesta.

Por exemplo, alguém que tenha apostado no cavalo favorito em cada um dos 144 mil páreos de corrida que tiveram lugar no Reino Unido nos vinte anos anteriores a 2010 veria seu cavalo ganhar mais ou menos 1 em cada 3 corridas. Isso parece muito impressionante, e produziria um lucro também impressionante se as proporções pagas para os favoritos fossem significativamente maiores que 2 para 1. Mas não são: as casas de apostas, caracteristicamente, oferecem prêmios menores para os favoritos. Como resultado, ao mesmo tempo que você pode ganhar cerca de $\frac{1}{3}$ das apostas, as perdas geradas pelos outros $\frac{2}{3}$ acabarão comendo todos os seus ganhos, e mais. Os registros mostram, na verdade, que, se você tivesse apostado £10 em cada favorito ao longo desses vinte anos, teria acabado com uma perda líquida bem maior que £100 mil.

Em contraste, a regra áurea das apostas mostra que há um jeito de ganhar dinheiro como apostador regular. Da mesma maneira que nos cassinos, ele envolve uma habilidade, e nesse caso é identificar onde as casas de apostas fizeram bobagem e ofereceram prêmios melhores do que deveriam. Mas não podem ser prêmios apenas ligeiramente melhores; o tamanho da bobagem deve ser grande o suficiente para incluir alguma margem de erro no julgamento, mais uma margem de lucro. Por exemplo, imagine que um cavalo no páreo das 14h30 em Ascot tenha uma chance decente de ganhar, e que as casas de apostas estejam oferecendo 3 para 1. A regra áurea diz que você só deve fazer a aposta se tiver confiança não só de que o cavalo tem uma “chance decente”, mas de que tem

uma probabilidade significativamente maior que a insinuada pelas chances da casa de apostas, ou seja, 25%. Acrescente uma margem de segurança, mais uma margem de lucro, e a regra áurea lhe diz que apostar nesse cavalo só faz sentido se as chances de ele ganhar forem de pelo menos 35%. Você realmente acredita que as casas de apostas erraram a esse ponto?

Essa é a pergunta que faz tropeçar a maioria dos aspirantes a jogador profissional. Eles acreditam que a pergunta-chave é simplesmente quem vai ganhar. Na sua busca de resposta, poderão passar horas e mais horas estudando resultados, publicações especializadas e sites on-line para formar uma imagem realmente detalhada de, digamos, algum time de futebol ou jogador de tênis – e identificar quando eles têm uma chance real. O jogador estrela do time voltou de uma contusão, digamos, ou o jogador de tênis tem bons resultados em quadras de saibro. Armado dessas informações, ele faz a aposta. Mas o que não levou em conta é que os especialistas da casa de apostas têm acesso à mesma informação e a muitas outras mais, e então fizeram o melhor para oferecer prêmios injustos para cada ganhador possível. Assim, os lucros obtidos em cada aposta ganha não compensam todas as apostas perdidas – garantindo que, a longo prazo, o apostador saia perdendo.

Para apostadores que deixam de fazer a pergunta certa, ganhos ocasionais – até mesmo frequentes – são perigosos, pois ajudam a mascarar as consequências a longo prazo. Só à medida que semanas, meses e anos vão passando é que se torna claro que os ganhos não se transformaram em lucros regulares. Eles estão sendo destruídos, de forma lenta mas firme, pela lei das médias.

Apostadores bem-sucedidos em esportes, como Veitch, conseguem seus resultados radicalmente diferentes adotando uma abordagem radicalmente diferente. Seu foco não está em identificar vencedores, mas em encontrar resultados cujas chances foram significativamente subestimadas pelas casas de apostas. Isso pode levá-los a agir de uma forma que espante os amadores, como apostar em diversos cavalos na mesma corrida. Se o seu foco está em achar vencedores, isso não faz sentido, pois só pode haver um vencedor. Mas para

aqueles que sabem encontrar apostas de valor, esta é a chave; então, é inteiramente possível enxergar vários exemplos no mesmo páreo.

Proceder assim, porém, é uma questão bem diferente – e alguns desconfiam que seja algo totalmente impossível. Muito tempo atrás havia oportunidades de sobra para os apostadores em determinados esportes ganharem dinheiro. Enquanto as casas de apostas focalizavam sua atenção nas ligas principais de esportes populares, jogadores especializados podiam varrer as chances divulgadas para jogos em partidas de ligas inferiores ou em esportes menos conhecidos, e descobrir apostas mal alocadas. Ainda que os prêmios fossem bastante modestos, esse era um trabalho duro. Contudo, desde meados dos anos 2000, não está claro que qualquer volume de trabalho árduo possa fazer fortunas nas apostas esportivas. Agora todas as grandes casas baseiam seus critérios em sofisticada análise estatística de dados passados, combinada com modelos computadorizados elaborados por consultorias especializadas. Além disso, fazem uso extensivo do cálculo de probabilidades produzido por bolsas de apostas, como a Betfair, que se baseia em informações de milhares de indivíduos, e pelos colossais mercados de apostas asiáticos. As proporções daí resultantes são produto da “sabedoria da multidão” e tornaram-se excepcionalmente confiáveis. Isto é, em milhares de eventos esportivos, aqueles com resultados que apresentam chances de 3 para 1, por exemplo, realmente acontecem 25% das vezes. Como as bolsas de apostas ganham dinheiro com um modelo de negócios radicalmente diferente daquele das casas de apostas (ou seja, pegando uma porcentagem de apostas vencedoras), as chances ali divulgadas são realmente estimativas de probabilidades reais, e não chances enganosas, deliberadamente rebaixadas para dar uma margem de lucro. Tudo isso significa que nunca foi mais difícil achar furos de casas de apostas – e fazer apostas de valor. Como diria um economista, o mercado de apostas esportivas nunca foi mais “eficiente”, com as chances divulgadas refletindo essencialmente toda a informação acessível para qualquer pessoa. O advento dos chamados *bots* de apostas – algoritmos de computador que detectam quaisquer chances

deslocadas nas bolsas de apostas – tem provocado o sumiço até das ineficiências temporárias.

Mesmo assim, pode haver oportunidades de ganhar um pouco de dinheiro em apostas esportivas para aqueles que estão dispostos a empenhar algum esforço nisso. O artifício é analisar dados passados, procurando fatores que as casas de apostas tenham deixado passar, criando ineficiências e, portanto, apostas de valor com chances obviamente generosas. Um desses fatores é o número de competidores nas corridas de cavalos. Um “campo” grande é mais desafiador para as casas de apostas em termos de estabelecer as chances de forma acurada, desconsiderando “desconhecidos” com boas chances – mas pode fazer também com que zebras atrapalhem os melhores corredores. De outro lado, campos pequenos são mais fáceis de avaliar e oferecem menos oportunidades de surpresa. Em alguma região intermediária – por exemplo, páreos envolvendo entre seis e dez competidores – existe uma oportunidade potencial para identificar apostas de valor.

Outro caminho é se concentrar nos mercados “de novidades”, tais como quantos chutes um time acerta na meta. As casas de apostas investem relativamente pouco em analisar esses mercados, e podem desprezar fatores que levam a apostas de valor. Qualquer que seja o caminho escolhido, encontrar e validar esses fatores envolve “garimpo de dados”, e, como veremos adiante, isso encerra armadilhas para os descuidados. Com toda a certeza, aqueles que têm êxito não ficam se gabando de como o conseguiram. Por isso, uma vez que se tornem amplamente conhecidos, esses fatores passarão a ser considerados no estabelecimento das chances divulgadas – destruindo qualquer valor que pudessem conter. Como diz Nick Mordin, analista britânico do sistema de corridas de cavalos e que dá palpites para apostas: “Os sistemas de apostas são como vampiros: quando você os arrasta para a luz do dia, eles morrem.”

Existe algum jeito mais fácil de ganhar dinheiro com apostas? Sim, a se acreditar nos argumentos mencionados por incontáveis sites na internet anunciando livros, programas de computador e serviços de dicas em tese capazes de identificar ganhadores. Será que funcionam? Muitos de fato identificam um

bom número de ganhadores, mas, como mostra a lei áurea das apostas, isso não é especialmente difícil, e tampouco é o ponto em discussão. A única maneira (legal) de ter lucro a longo prazo com as apostas é identificando apostas de valor. Alguns dos serviços especializados em dar dicas podem alegar que fazem isso, no entanto, aí, o problema é a ganância. Uma vez que um serviço de dicas se mostre confiável, inevitavelmente atrairá aqueles que têm mais dinheiro que bom senso, e que tentam pôr quantias maciças de dinheiro nas casas de apostas. Sempre alertas para novas ameaças, elas reagirão reduzindo a proporção do prêmio para proteger sua margem – destruindo assim qualquer valor. E isso se as casas de apostas efetivamente aceitarem o seu dinheiro. Para apostadores profissionais sérios como Veitch, identificar apostas de valor é uma coisa; ser capaz de explorá-las com dinheiro sério é outra.

As casas de apostas, ansiosas para proteger seu fluxo de caixa, podem se recusar, e de fato se recusam, a aceitar apostas daqueles que elas julgam realmente saber o que estão fazendo, e “conviver” com essa situação passa a ser um grande desafio. Agências de apostas on-line possuem programas que identificam apostadores cujo sucesso ameaça seus modelos de negócios, e “impõem às suas apostas limites” ridiculamente baixos – ou simplesmente fecham as contas desses jogadores.

A maioria das pessoas que “ficam empolgadas” aposta apenas por diversão – talvez uma vez por ano, num grande evento como o Grande Prêmio Nacional de turfe, no Reino Unido, ou o Superbowl ou o Kentucky Derby, nos Estados Unidos. Nunca pensam em jogar como meio de ganhar a vida. Isso é ótimo, pois a maioria das pessoas não tem consciência da regra áurea das apostas, muito menos de suas implicações para apostas bem-sucedidas. A realidade é que, como ocorre nas apostas dos cassinos, a menos que você esteja disposto a investir muito esforço, o jeito mais provável de fazer uma pequena fortuna com apostas é já começar com uma fortuna grande.

Conclusão

É inteiramente possível ser um apostador bem-sucedido. Isso só requer três coisas: compreensão da regra áurea das apostas, perícia para encontrar oportunidades que sejam coerentes com ela e um temperamento capaz de lidar com os caprichos do acaso. A evidência sugere que pelo menos 95% de nós simplesmente não têm o que é necessário.

ⁱ O pano amarelo é lançado pelos juízes auxiliares com o intuito de chamar a atenção do juiz principal para a ocorrência de alguma falta ou irregularidade na jogada. (N.T.)

16. Garantir – ou arriscar?

QUER GOSTEMOS, QUER NÃO, todos nós temos de fazer apostas. Elas podem não envolver um cassino ou uma casa de apostas, mas ainda assim implicam dinheiro e incerteza. Se você possui um imóvel, terá um seguro para ele, e provavelmente também para seu conteúdo. Em outras palavras, você despende uma quantia considerável refletindo sua visão acerca de um evento incerto: alguma calamidade atingindo sua casa. Isso é uma aposta – bem como o seguro-saúde, o seguro de vida e os investimentos. Mas são boas apostas? Essa é uma pergunta que provavelmente cruza a mente da maioria das pessoas que compraram produtos eletrônicos de consumo e lhes foi oferecida uma “garantia estendida”. Houve tempos em que ela só era ofertada para itens caros, mas, a partir de meados dos anos 1990, passou a ser apresentada para quase tudo, de telefones a frigideiras. E hoje ainda é um grande negócio: só no Reino Unido, milhões de pessoas aceitam anualmente a oferta de garantia estendida, gastando cerca de £1 bilhão em apólices. Contudo, também tem havido muita controvérsia sobre se ela vale a pena. Alguns insistem em que a taxa de defeitos da maioria dos produtos eletrônicos é baixa demais para justificar a cobrança dessas apólices. Outros argumentam que tudo é um pouco mais complicado que apenas uma questão de probabilidades: aqueles que pagam a garantia estendida estão comprando paz de espírito, bem como a cobertura para substituição do produto.

A verdade é que as apostas da vida real são bem mais sutis que aquelas feitas em cassinos e similares. Felizmente, os conceitos básicos para compreendê-las foram desenvolvidos séculos atrás. O resultado é uma das aplicações mais úteis das leis da probabilidade – e também uma das mais controversas. O ponto de partida para toda decisão tomada diante da incerteza é uma pergunta: quais são as prováveis consequências? O método básico para respondê-la foi desenvolvido pelo brilhante polímata francês do século XVII, pioneiro da teoria da

probabilidade: Blaise Pascal. E para algo tão poderoso, é impressionantemente simples: as consequências que devemos esperar de um evento incerto podem ser avaliadas multiplicando-se essas consequências pelas chances de o evento realmente ocorrer.

Suponha, por exemplo, que nos seja oferecida uma aposta com 20% de chance de ganhar £100. As £100 são a consequência de a aposta sair a nosso favor, então, de acordo com o argumento de Pascal, as consequências que devemos esperar obter desse evento incerto são £100 vezes 20% de chance de ocorrer, dando um valor esperado de £20. Tudo muito simples – mas será que faz sentido? Afinal, jamais ganharíamos na realidade um prêmio de £20; ganharíamos £100 ou nada. É verdade, você só fica sabendo depois de ter feito a aposta, e aí já é um pouco tarde. A beleza da regra de Pascal é que ela nos permite estimar quanto a aposta vale a pena *antes de efetivamente fazê-la*. Para isso, imagine que, no decorrer da sua vida, você tenha enfrentado uma grande quantidade dessas apostas de “1 chance em 5” – tão grande que a lei dos grandes números seja bastante confiável. Sabemos que ganharíamos aproximadamente 20% todas as vezes. Em média, levaríamos para casa 20% de todos esses prêmios de £100 que nos foram oferecidos. A regra de Pascal simplesmente aplica o mesmo raciocínio para cada aposta individual. E, ao fazê-lo, ela nos dá um número – o *valor esperado* – que nos permite decidir se alguma aposta vale a pena ser feita, antes da hora. Só temos de perguntar a nós mesmos se o valor esperado para ganhar vale o custo da participação.

No caso da nossa aposta de 20%, calculamos que o valor esperado para ganhar são 20% de £100, ou £20. Mas não devemos cair na armadilha de tantos amadores e nos deixar hipnotizar pela perspectiva de ganhar; devemos também encarar a possibilidade de perder – e há 80% de risco de que isso ocorra. Então, vamos aplicar novamente a regra de Pascal, dessa vez para as nossas perdas. Claramente, não queremos que as perdas esperadas excedam os ganhos esperados, porque isso significa que perderemos dinheiro a longo prazo. No nosso exemplo, podemos fazer isso garantindo não arriscar tanto dinheiro que perder 80% dele exceda os ganhos esperados, que já sabemos ser de £20. Assim,

não devemos arriscar mais que £25 (pois 80% disso equivale a £20). Claro que você pode se dar bem fazendo isso uma vez, até mesmo algumas vezes, mas continue quebrando a regra de Pascal, e você vai acabar se lamentando.

O poder da regra de Pascal pode ser aplicado a mais que joguinhos tolos. Para apostadores profissionais, ela é a luz que os guia rumo ao dinheiro sério, e sustenta a lei áurea das apostas. Nesse caso, estamos tentando saber se a recompensa (na forma das proporções oferecidas) é razoável, dado nosso julgamento sobre as chances de ganhar. Se o valor esperado de um ganho exceder o custo esperado de uma perda numa margem confortável, então teremos obtido uma “aposta de valor”. Valores esperados também são cruciais para avaliar as “apostas” com que deparamos ao jogar nesse grande cassino cósmico que chamamos de vida.

Tomemos o caso das garantias estendidas. Em 2013, a revista da Associação de Consumidores do Reino Unido, *Which?*, examinou o que chamou de “a grande exploração das garantias estendidas”. A investigação da revista centrava-se no fato de as lojas darem informação imprecisa sobre as garantias, e concluía que elas não valiam o dinheiro pago, o que provavelmente não é surpresa para muita gente. Todavia, mesmo fazendo algumas afirmações rebuscadas para respaldar sua conclusão, a revista fracassou em mostrar o tamanho exato da exploração das garantias. Fazer isso é um bom exercício sobre a utilidade do valor esperado. No levantamento da *Which?*, descobriu-se um supermercado cobrando £99 pela cobertura de cinco anos para uma TV que valia £349 quando nova. Agora, se você acabou de despendar essa quantia, £99 pode não parecer muito a se pagar para ter mais cinco anos de paz de espírito. No entanto, a aplicação da teoria do valor esperado pode levá-lo a parar para pensar. Se a TV quebrar, a “perda esperada” é o custo da TV multiplicado pelas chances de quebra – o que não sabemos. O que sabemos, porém, é que esse valor esperado não deve ser maior que o pagamento da garantia de £99 que estão nos pedindo – porque então estaremos pagando por um risco maior que as chances de a TV quebrar. Isso nos diz que a garantia só vale a pena se for menos que £349 multiplicado pelo risco de quebra, ou que o risco de a TV quebrar durante os

cinco anos precisa ser de *pelo menos* $99/349 = 28\%$. Se você acha isso razoável, então vá em frente, mas talvez queira checar qual a real taxa de estrago – como a *Which?* fez. A taxa real de quebras é de apenas 5%, o que está muito abaixo da taxa de quebra mínima para que o pagamento de £99 seja justo. E também nos permite calcular qual deveria ser um pagamento justo: £349 vezes a taxa de quebras de 5%, ou cerca de £18 – apenas uma fração do valor de £99 que é cobrado.

A garantia da TV não foi sequer o pior caso: uma rede de lojas de eletrodomésticos oferecia uma garantia *premier* de cinco anos por £139 para uma TV que custava £269 – o que já soa ridículo, mesmo sem fazer os cálculos matemáticos. No entanto, considerando que a taxa de quebra era de apenas 2%, o pagamento justo seria 26 vezes menor do que aquilo que estava sendo cobrado. Com margens de lucro dessas, não é surpresa que a *Which?* tenha encontrado tantos varejistas ávidos para nos empurrar garantias estendidas – ou pelo menos para quem não sabe fazer os cálculos. Agora nós sabemos, graças à regra de Pascal do valor esperado. Essa regra mostra que nos cobrarão mais pelo seguro se o prêmio exceder em muito o valor do produto multiplicado pela chance de ele quebrar durante o período segurado. Pelo menos no caso da TV, a taxa de quebra é de poucos por cento, então o pagamento justo não deveria ser mais que um baixo percentual do preço de aquisição (e isso ainda ignora a depreciação).

A mesma ideia básica pode ser usada quando se faz um seguro de perda ou roubo de um aparelho: um prêmio razoável a se pagar é aproximadamente o valor do aparelho vezes a chance de o evento ocorrer. Aqui vale a pena checar as estatísticas criminais, pois elas frequentemente revelam que um prêmio de mais de poucos por cento do valor do aparelho é uma completa exploração. Na ausência de estatísticas sólidas, a experiência pessoal pode ajudar a estimar os riscos. O simples fato de algo *não* ter acontecido com você já constitui uma informação surpreendente. Um pouco de matemática mostra que, se um evento nunca ocorreu, apesar de ter N oportunidades de ocorrer, então pode-se ter uma boa confiança de que a frequência de ele acontecer não é mais que 3 dividido por N . Assim, por exemplo, se, ao longo dos últimos cinco anos, você nunca perdeu

qualquer objeto que possua em circunstâncias similares às aquelas nas quais planeja usar seu novo aparelho, as chances de este ser o primeiro são provavelmente menores do que cerca de $3/N$, onde N é o número de suas posses relevantes. Se você imagina que tem mais que algumas dezenas de objetos desse tipo, então $3/N$ é cerca de 10%, e um prêmio justo a ser pago por cinco anos não seria mais que 10% do preço, dando um prêmio anual de cerca de 2% do preço de aquisição.

Algumas pessoas (especialmente aquelas que trabalham no ramo de seguros) irão protestar dizendo que tudo isso é simplista. E, sob alguns aspectos, é. Nós ignoramos o fato de que o seguro com frequência fornece mais do que apenas o custo da substituição; muitas políticas incluem serviços como assistência técnica de 24 horas no local. Vale a pena pagar pela paz de espírito e a conveniência, mesmo que elas sejam difíceis de quantificar. Então, existe o problema de ser capaz de lidar com as consequências se sua “aposta” relativa à necessidade de seguro der errado – o que, como estamos lidando com eventos incertos, é sempre possível. Tem todo cabimento recusar um seguro que exija dez vezes aquilo que você encara como prêmio justo se você puder lidar com as consequências, caso sua decisão inteiramente racional se mostre errada. Se for uma engenhoca qualquer ou, digamos, uma máquina de lavar, o custo pode ser um aborrecimento, mas não catastrófico.

Algo muito diferente é o seguro de sua casa, digamos, ou da cobertura médica em viagens para o exterior. Você pode muito bem pensar que o risco de ficar doente numa viagem curta é tão baixo que pagar (digamos) um prêmio de £20 não vale a pena. Mas, com as contas hospitalares e os custos de repatriação capazes de exceder 10 mil vezes essa quantia, será que você está *realmente* apostando com confiança que as chances de ocorrer algum problema de saúde sejam mesmo menores que 1 em 10 mil, quando o custo de perder a aposta são assustadoras £200 mil?

Isso ressalta um fato-chave acerca dos seguros e, na verdade, sobre tomadas de decisão em geral: o contexto é tudo. Se você é pobre, mesmo um prêmio justo pode estar além das suas possibilidades; independentemente de quão racional

você seja, não tem escolha a não ser confiar na sorte. Por outro lado, gente rica pode estar disposta a pagar mais que o prêmio justo simplesmente porque, para elas, o dinheiro significa menos. O fato é que, enquanto a porcentagem pode ser a mesma, um multimilionário que pagar £10 milhões em £100 milhões não sentirá tanta dor quanto alguém que vive de pensão pagar £10 em £100.

Essa dependência do valor do dinheiro em relação ao contexto é crucial para se tomar decisões sobre ele, e foi percebida pelos pioneiros da teoria da probabilidade. No começo do século XVIII, a regra de Pascal para tomadas de decisão utilizando o valor esperado havia se tornado amplamente conhecida. Ela parecia dizer que todas as decisões envolvendo dinheiro podiam ser tomadas multiplicando a probabilidade de cada resultado pela quantia implicada. Mas em 1713 o matemático suíço Nicolau Bernoulli (cujo tio era Jacob Bernoulli, famoso pelo teorema áureo) apontou um problema. Em termos simples, indicou que a regra de Pascal podia levar as pessoas a tomar decisões absolutamente irrealistas.

Por exemplo, imagine que você é convidado a participar de um jogo em que a moeda é lançada mil vezes, e você ganha quando sair cara – e, para tornar as coisas um pouco mais interessantes, o prêmio dobra a cada lance, até finalmente sair cara. Quanto você deveria estar disposto a pagar para participar disso? A regra de Pascal diz que o “valor esperado” de se jogar é a probabilidade de ganhar um lançamento – ou seja, 50% – multiplicada pela quantia em oferta, que dobra a cada lance. Claro que, quanto mais o jogo continua, maior é o prêmio, contudo, também maiores são as chances de o jogo parar. Aplicando a regra de Pascal, esses dois efeitos que se contrabalançam levam a um valor esperado de... infinito. A decisão é clara: você deveria vender tudo que tem para jogar, e os ganhos esperados são infinitos. Todavia, conforme mostrou Bernoulli, isso é ridículo. Primeiro, as chances de recuperar a taxa de participação infinita nos ganhos são essencialmente nulas. Para ganhar até a modesta soma de £16, era preciso que a moeda fosse lançada quatro vezes antes de dar cara, e há apenas 3% de chance de isso acontecer. Depois, há o pequeno problema de que, de qualquer maneira, os organizadores do jogo só teriam uma soma finita para nos

pagar. Contudo, a regra de Pascal nos diz que ainda assim faz sentido ignorar tudo e pagar uma quantia infinita.

Esse resultado bizarro veio a ser conhecido como paradoxo de São Petersburgo, pois a pessoa que o resolveu (Daniel Bernoulli, primo de Nicolau) revelou sua solução para a Academia de Ciências desta cidade em 1783. Ao mesmo tempo que o problema em si parecia um daqueles jogos mentais bobos que os estudiosos adoram, ele levou Bernoulli a inventar um conceito que hoje apoia o ramo de seguros, um negócio global de US\$ 100 bilhões: utilidade. Ao fazê-lo, ele estabeleceu uma surpreendente conexão entre o frio e platônico mundo da matemática e o caloroso e impreciso mundo da psicologia humana.

Segundo Bernoulli, o paradoxo só existe porque a regra de Pascal se concentrava somente na aritmética, deixando de considerar a noção subjetiva do *valor* do dinheiro. Este, argumentou Bernoulli, depende do contexto – em especial, em quanto temos dessa coisa. Um bilionário vê menos valor – ou “utilidade”, no jargão – em £100 mil que alguém que vive da previdência social. Contudo, mesmo o bilionário pode enxergar alguma utilidade nele. Bernoulli, portanto, propôs que, ao tomar decisões envolvendo dinheiro, a regra de Pascal deveria nos dar não o valor esperado das consequências, mas a *utilidade* esperada. O que é ela – e como a calculamos para uma dada soma de dinheiro?

Bernoulli deduziu uma regra de conversão simples, com base no seu argumento de que dinheiro extra sempre acrescenta *alguma* utilidade extra, e o efeito se dilui à medida que ganhamos mais. Matematicamente, isso implica que a utilidade de uma quantia é proporcional ao seu logaritmo. Assim, por exemplo, £1 mil tem uma utilidade de 3 unidades porque o logaritmo de 1 000 é 3, enquanto £1 milhão – uma quantia mil vezes maior – tem apenas 3 unidades de utilidade extra, pois o logaritmo de 1 milhão é 6. Isso, argumentou Bernoulli, tem um grande impacto sobre como as pessoas com diferentes níveis de riqueza encaram as decisões monetárias. Visto sob o prisma da utilidade, se você tem £1 mil e recebe uma oferta de ganhar outras £1 mil, isso equivale a um salto em utilidade de 3 para 3,3 *unidades*, pois o log de 2 mil é 3,3. Em contraste, alguém com £1 milhão já tem uma utilidade de 6, e ganhar £1 mil significa aumentar a

utilidade financeira para o log de 1 001 000 – ou seja, um aumento de 6 para 6,0004, o que dificilmente valeria a pena.

Ainda que se possa debater a “taxa de conversão” precisa de dinheiro para utilidade, um ponto-chave é que eles não mudam em proporção direta: a utilidade cresce mais devagar com o aumento da riqueza. E é isso que pode transformar o “tolo” paradoxo de São Petersburgo em algo sensato. Se usarmos a regra de Pascal e trocarmos os ganhos esperados pela *utilidade* esperada de participar, o resultado é drástico. À medida que aumenta o número de lançamentos, a utilidade esperada não vai ficando cada vez maior, mas, em vez disso, estabiliza-se num valor finito sensato – e claramente não há sentido em pagar uma quantia infinita para ganhar isso.

Com o correr dos anos, os estudiosos vêm discutindo os méritos da resolução do paradoxo apresentada por Bernoulli e suas lições para tomar decisões na vida real. É difícil imaginar alguém idiota o suficiente para pagar uma soma de dinheiro infinita em troca de qualquer coisa (difícil, mas não impossível – veja o Box a seguir). O que não está em dúvida é o efeito transformador que tudo isso tem no teimoso ramo dos seguros – como o próprio Bernoulli percebeu. Oferecer-se para compensar as pessoas pelas suas desgraças é uma ideia adorável. Mas precisa fazer algum sentido financeiro. Uma vez que estamos lidando com eventos incertos, esse não é um problema trivial. Para começar, as seguradoras querem garantir que haja dinheiro suficiente entrando via prêmios das apólices para cobrir as desgraças. Isso significa estimar o risco provável e estabelecer prêmios um pouco maiores que o estimado golpe financeiro de se custear uma calamidade, usando a mesma lógica que leva as casas de apostas a oferecer pagamentos inferiores ao que as chances sugerem ser justos. Fazer seguro de muita gente também ajuda, pois a lei das médias aproximará a frequência das desgraças da taxa esperada – pelo menos em teoria.

O conceito de utilidade de Bernoulli leva a muitas outras consequências mais sutis, como mostrar que as seguradoras que dividem riscos com as concorrentes podem ao mesmo tempo reduzir sua própria exposição e oferecer prêmios mais baixos a seus clientes – todo mundo sai ganhando. Em termos simples, a teoria

de Bernoulli permite que elas cubram muitos riscos que de outra forma teriam recusado. Em geral, essa é uma coisa boa, permitindo-nos comprar paz de espírito para tudo, desde férias canceladas até aquecedores quebrados. Mas há quem a veja como um meio de explorar nossas neuroses. Sem dúvida é o que parecem as políticas de garantia estendida. Na realidade, as seguradoras sabem que o conceito de utilidade de Bernoulli tem apenas uma – com o perdão do jogo de palavras – utilidade limitada. Se um risco é muito alto ou vago, a diferença entre o prêmio justo e aquilo pago pelos clientes torna-se pequena demais para valer a pena comercialmente. Com frequência a cobertura de riscos públicos cai nessa categoria: são riscos difíceis de julgar, e as indenizações podem ser colossais. Isso levou a indústria de seguros a desenvolver uma variedade de técnicas que lhe propiciam a cobertura contra as vicissitudes da vida. Algumas são bastante simples – tais como apenas cobrir perdas acima de certo mínimo, ou “excesso”. Outras são produto de detalhados cálculos de probabilidade que permitem às seguradoras assumir riscos realmente extraordinários, tais como a teoria dos valores extremos – que iremos ver em capítulo posterior.

Como os cassinos, as companhias de seguro construíram seu modelo de negócios sobre as leis da probabilidade – e ela funciona bem para elas. Na maior parte do tempo, também funciona bem para nós, embora às vezes desconfiemos que estamos sendo explorados. No entanto, não precisamos fazer uma escolha rígida entre eliminar o seguro ou simplesmente arriscá-lo. Há um meio-termo – pelo menos para itens pequenos, que por acaso é onde se encontra a maioria dos abusos. A regra de Pascal nos permite sermos nossos próprios seguradores. Simplesmente calculamos um prêmio justo multiplicando o valor do item pelo risco de sinistro, e então pagamos esse valor em parcelas para o nosso próprio fundo de amortização. Como alternativa, podemos economizar o prêmio que, de outra maneira, teria ido para a seguradora – e pode ter certeza de que será mais que suficiente. De qualquer maneira, estamos cobertos contra sinistros, ou, se eles não ocorrerem, acabaremos com uma bela poupança.

Em 1957, David Durand, professor de finanças no Instituto de Tecnologia de Massachusetts (MIT), apontou alguns paralelos perturbadores entre o jogo “absurdo” atacado por Bernoulli e o investimento nas chamadas ações de crescimento (*growth stocks*). Essas ações são de empresas cujas receitas parecem estar estourando nas alturas. As empresas habitualmente dão manchetes na mídia, despertando enorme interesse nos investidores. Enquanto muitos entram na dança de qualquer maneira, os investidores sérios preferem sondar mais um pouco, para descobrir se o preço da ação é justificado pelas perspectivas da companhia. Em termos simples, isso envolve estimar o valor presente da performance e do ativo futuros da empresa, assumindo certas taxas de crescimento e taxas de juros. O problema, claro, é que ninguém sabe ao certo quais serão essas taxas futuras. Pior ainda, o chamado processo de “desconto” pressupõe que a empresa continue a existir para sempre – como o jogo que está no cerne do paradoxo de São Petersburgo. Analistas financeiros que pressupõem taxas de crescimento nunca abaixo das taxas de juros acabam com avaliações consistentes com preços de ações iguais a... infinito. Decerto ninguém seria idiota a ponto de acreditar numa “análise” dessas. Pense outra vez. Um estudo publicado em 2004 pelos matemáticos Gabor Székely e Donald Richards concluía que um fenômeno do tipo do paradoxo de São Petersburgo foi o fator-chave da conhecida “Bolha da Internet”, do fim da década de 1990. As “ações de crescimento” eram empresas de alta tecnologia que nunca tinham dado lucro, mas cujas ações subiam a níveis estratosféricos – mas consistentes com as avaliações malucas. Quando estourou, a Bolha da Internet varreu US\$ 5 trilhões do Nasdaq, o mercado de ações americano, onde as ações eram negociadas. Ainda assim, na verdade devemos nos considerar sortudos; poderia ter sido mais – na realidade, infinitamente mais.

Claro que faz sentido ser pessimista. É possível que muitas calamidades nos atinjam de uma só vez antes que os prêmios tenham sido pagos, então, deve-se pôr uma quantia decente no fundo de amortização para cobrir essa possibilidade. Também não devemos nunca perder de vista o propósito desse fundo: ele está lá para ser usado se, e somente se, houver alguma calamidade. Evidente que nos sentiremos realmente irritados se nosso brilhante estratagema não der certo, mas isso é algo com que temos de conviver. Como veremos no próximo capítulo, tomar decisões sobre riscos nem sempre é racional. Como lidam com a probabilidade, a regra de Pascal e a teoria da utilidade não podem dar garantias – e nós devemos ter uma provisão para o caso de falha do plano mais bem-elaborado. No entanto, como meio de manter dinheiro no bolso, em vez de dá-lo de presente, elas não têm preço.

Conclusão

Nós vivemos num mundo cheio de riscos, e o seguro foi inventado para nos ajudar a lidar com as consequências – ao mesmo tempo que damos lucro para as companhias de seguros. Regras práticas simples mostram quando o seguro não vale e quando é melhor fazer um – e como guardar uma provisão para quando até o melhor dos planos falha.

17. Fazer apostas melhores no cassino da vida

VALE A PENA PEDIR um aumento ao patrão? Devemos agir segundo os rumores de como o nosso bairro vai mudar? Qual a melhor maneira de lidar com o aquecimento global? Todo dia somos confrontados com tomadas de decisão, ou pelo menos com a necessidade de ter uma opinião sobre elas. Contudo, até as menos importantes muitas vezes parecem carregadas de pressão, com suas múltiplas incertezas e consequências. Combine isso com o medo de tomar uma decisão errada, e não é nenhuma surpresa que simplesmente decidamos não decidir. Por sorte, tomar decisões diante da incerteza tem sido há muito tempo uma grande parte da teoria da probabilidade, resultando numa gama de ferramentas capazes de dissecar a complexidade. Elas são notáveis pelo seu poder de extrair conclusões acerca de grandes questões com pouco esforço. O originador do que hoje se chama teoria da decisão, o brilhante polímata francês Blaise Pascal, usou-a para atacar uma das grandes questões definitivas: faz sentido acreditar em Deus?

Contrariamente ao que às vezes se alega, Pascal não estava tentando provar a existência de Deus. Em sua opinião, Deus era tão inefável e incompreensível que qualquer prova desse tipo – ou, na verdade, qualquer refutação – não significava muita coisa. Uma pergunta que valia a pena ser feita, argumentava Pascal, era se tinha cabimento acreditar em Deus. Ele começou por retornar ao seu conceito de expectativa, segundo o qual não são apenas as probabilidades dos resultados que importam, mas suas correspondentes consequências. Quanto ao que são essas consequências, Pascal era bastante vago, mas a essência delas pode ser resumida na seguinte tabela:

	DEUS EXISTE	DEUS NÃO EXISTE

OPÇÃO POR ACREDITAR	<i>Consequência:</i> positiva – eternidade no paraíso	<i>Consequência:</i> negativa –perda de tempo e esforço em rituais
OPÇÃO POR NÃO ACREDITAR	<i>Consequência:</i> negativa – potencialmente grande encrenca por causa de um Deus rancoroso	<i>Consequência:</i> positiva – economia de tempo e esforço em rituais

Por que a crença em Deus faz sentido, de acordo com Blaise Pascal.

Note que, em contraste com uma aposta simples, não deparamos mais com um resultado direto tipo “ganhar/perder”. Em vez de ter de decidir se Deus existe ou não, Pascal mostra que é possível lidar com situações mais complexas, envolvendo ambas as possibilidades. Como a tabela mostra, há agora quatro cenários com os quais lidar. Para decidir qual é a melhor opção, Pascal sugeriu que elaboremos as consequências *esperadas* de cada uma. Isso significa multiplicar cada consequência pela respectiva probabilidade. Mas como devemos estimar as chances da existência de Deus? Aparentemente argumentando que não havia como a razão preferir uma alternativa à outra, Pascal optou por estabelecê-las como iguais: 50:50. Você não precisa ser ateu para pensar que há algo de errado nisso; afinal, se você não soubesse nada sobre um cavalo, assumiria sem mais problemas que ele teria uma chance parelha de ganhar o páreo? Pascal estava lutando com uma dificuldade que até hoje causa controvérsias: que probabilidade atribuir a uma coisa sobre a qual você não sabe nada. Voltaremos a encontrar isso em outros contextos, mas por enquanto vamos apenas seguir em frente – já que, de todo modo, Pascal está prestes a usar um truque que causa um curto-circuito em todo o problema.

Admitindo por enquanto que existe realmente uma chance igual de Deus existir ou não existir, as probabilidades são as mesmas em todos os casos, e seu impacto se cancela. Somos deixados com a simples comparação das consequências de cada escolha, para verificar qual delas é a melhor. Segundo a coluna da direita na tabela de Pascal, o melhor resultado oferecido, se Deus não existir, é apenas economizar tempo e esforço. Por sua vez, o melhor resultado oferecido, se Deus existir, é a eternidade no paraíso. Segundo esse raciocínio, a crença em Deus faz perfeito sentido. Pelo menos segundo a forma que Pascal

montou; mas e se não aceitarmos seu argumento de uma chance de 50:50 de Deus existir? Agora temos de elaborar as quatro consequências *esperadas* na sua totalidade, multiplicando cada consequência individual pela respectiva probabilidade e vendo que combinação é a melhor. Tudo muito tedioso e problemático. No entanto, como matemático, Pascal sabia que havia um meio de evitar tudo isso. Ele declarou que as consequências de acreditar num Deus que de fato existe – isto é, vida *eterna* no paraíso – não são meramente positivas, são *infinitas*. Como todas as outras consequências são meramente finitas, não importa quais sejam as várias possibilidades: a decisão que leva à única gratificação infinita ganha. Num só golpe, Pascal fez a crença em Deus ser a única decisão racional.

Mais uma vez, se acha que tudo isso é altamente suspeito, você está em boa companhia; hoje, poucos estudiosos levam o argumento de Pascal a sério, por todas as razões apresentadas e outras mais. O que devemos todos considerar, porém, é sua abordagem básica para decidir entre várias opções. Usada com menos artifício, ela pode dissecar muita complexidade e nos dar decisões claras e definidas em face da incerteza. Nem precisamos ir tão longe a ponto de fazer somas; simplesmente escrever uma tabela como aquela da aposta de Pascal muitas vezes ajuda a iluminar o melhor curso de ação.

Suponha que uma fábrica está tentando resolver como reagir à notícia de que um produto químico que vinha sendo usado pode ser ruim para o ambiente. O problema é que a evidência não é muito convincente, e talvez não consiga passar pelo teste do tempo. Assim, a empresa se defronta com uma tomada de decisão em situação de incerteza. Logo, vamos criar uma tabela das várias consequências.

Não vai ser fácil transformar essas consequências em números e multiplicá-los pela probabilidade desconhecida de o produto químico *realmente* ser tóxico. Então vamos tirar uma folha do caderno de Pascal e ver se isso pode ser evitado. Não precisamos ir tão longe quanto ele foi, ao trazer o infinito para a mistura. Em vez disso, vamos simplesmente procurar aquilo que se chama “dominância” – isto é, ver se uma decisão é melhor *independentemente* das probabilidades. No

exemplo, é óbvio que, se o produto químico realmente for tóxico, trocar para um substituto é a melhor decisão. Mais traiçoeiro é escolher entre as consequências de cada decisão se o produto *não* se provar tóxico. Contudo, se considerarmos que o movimento e o custo da mudança não são grandes demais e poderiam ser facilmente justificados pelos consequentes benefícios para a imagem, então fica evidente que trocar ainda é a melhor coisa a se fazer. Como essa solução nos dá as melhores consequências *independentemente* de que o produto seja tóxico ou não, não precisamos mais nos preocupar em estabelecer a probabilidade exata: trocar para um substituto é sempre melhor.

	O PRODUTO É TÓXICO	O PRODUTO NÃO É TÓXICO
DECISÃO A: CONTINUAR USANDO O PRODUTO QUÍMICO	<i>Consequência:</i> prejudicial em termos ambientais; processos judiciais, má publicidade.	<i>Consequência:</i> os negócios continuam como sempre, mas a empresa pode parecer negligente.
DECISÃO B: TROCAR POR UM SUBSTITUTO	<i>Consequência:</i> bom para o ambiente, bom para a imagem da empresa.	<i>Consequência:</i> mudanças desnecessárias, mas a empresa daria impressão de responsabilidade.

Como a empresa deve responder a reclamações sobre seu produto?

Cada caso deve ser considerado com seus próprios méritos, claro, mas há o fato de que às vezes existe uma estratégia dominante que possibilita chegar à melhor decisão sem ficar se preocupando com as chances envolvidas. Habitualmente, porém, temos de trazer alguns números para captar os méritos relativos das várias consequências. Não importa qual seja a amplitude: -10 a +10, do pior para o melhor, é uma amplitude tão boa quanto qualquer outra. Assim, por exemplo, uma família pode estar considerando mudar de casa depois de ouvir boatos de que uma estrada está prestes a ser construída nas proximidades, e, após discutir o assunto, veio com a seguinte análise e o escore relativo dos benefícios para as várias consequências.

Diferentemente do caso da empresa obrigada a lidar com a ameaça química, a família não pode optar por uma decisão do tipo tiro certo, que seja a melhor independentemente de os boatos serem verdadeiros ou não. Para tomar sua

resolução, precisa comparar as consequências *esperadas* de cada decisão, e isso requer alguma estimativa da probabilidade de que os boatos sejam verdadeiros. No entanto, mais uma vez, podemos contornar esse problema traiçoeiro. Agora, conquanto as probabilidades sejam necessárias para tomar uma decisão, não temos de especificá-las. Em vez disso, podemos inverter o problema e perguntar qual *precisa ser* a probabilidade para a mudança de casa fazer sentido. Um pouco de matemática simples¹ mostra que, nesse exemplo, mudar de casa faz sentido se a família acredita haver chance maior que 1 em 3 de os boatos serem verdadeiros. Se parecerem implausíveis, devem ficar onde estão.

	BOATOS VERDADEIROS	BOATOS FALSOS
DECISÃO A: FICAR NA CASA	<i>Consequência:</i> localização barulhenta e insegura, mais dificuldade de vender a casa. SCORE: -10	<i>Consequência:</i> as coisas continuam como estão. SCORE: +7
DECISÃO B: MUDAR DE CASA	<i>Consequência:</i> nenhuma ameaça de estrada, porém locomoção e viagens à escola mais demoradas. SCORE: +2	<i>Consequência:</i> inquietação e gastos desnecessários, mas talvez seja hora de mudar. SCORE: +1

Se tomar uma decisão tão importante baseado em números como esses faz com que você se sinta desconfortável, então considere a alternativa: usar seu instinto visceral. Este nos expõe ao perigo de tomar nossas decisões com base em fatores que têm impacto emocional, mas que na realidade são irrelevantes. Se você acha que está imune a tais fraquezas humanas, imagine-se como o obstinado diretor executivo de uma empresa com 450 funcionários passando por um período difícil. Você sabe que provavelmente terá de reduzir os negócios, então está enfrentando decisões que podem ter grande impacto sobre sua força de trabalho. Decidido a fazer o melhor pelos funcionários, você contrata uma das melhores firmas de consultoria administrativa para decidir sobre o rumo a tomar. De modo absolutamente tradicional, eles lhe entregam um robusto relatório, uma fatura condizente – mas nenhuma recomendação. Em vez disso, apresentam duas opções:

Plano A1: reestruturar a empresa salvando 150 empregos.

Plano A2: não fazer nada, o que inclui 2 chances em 3 de fechar; 1 chance em 3 de salvar os 450 empregos.

Então, qual deles você escolhe? Se for como a maioria das pessoas, opta pelo Plano A1, com a certeza relativa dos 150 empregos. Mas, consciente do enorme impacto de sua decisão sobre a força de trabalho, você pede a opinião de uma segunda firma de consultoria para ter certeza de que examinou todas as opções. O resultado é outro relatório robusto e outra fatura – e nada de recomendação. Porém, mais uma vez eles oferecem dois planos:

Plano B1: continuar normalmente, provocando a perda de 300 empregos.

Plano B2: reestruturar a empresa, com 1 chance em 3 de que todo mundo mantenha o emprego, mas 2 chances em 3 de que os 450 empregos desapareçam.

Agora, qual dos dois parece melhor? O Plano B1 parece realmente terrível, enquanto o Plano B2 parece contar com alguma esperança. Então, agora é só uma questão de decidir se você escolhe o Plano A1 ou o Plano B2 – mantendo os respectivos consultores administrativos para fazer a reestruturação. É isso mesmo? Enquanto você faz a comparação, pode começar a perceber uma coisa estranha: a promessa do Plano A1 de salvar 150 empregos na força de trabalho total de 450, via reestruturação, não é a mesma coisa que a lúgubre advertência do Plano B1, de que continuar como está levará à perda de 300 empregos? Aplicar um pouquinho da teoria de Pascal revela outro fato singular: a promessa do Plano B2, de 1 chance em 3 de que todos os 450 funcionários conservem seus empregos, implica uma perda esperada de $450 \times \frac{1}{3} = 150$. É exatamente o que o Plano A2 oferece. Em suma, em termos de dispensas prováveis, os planos são idênticos. A única diferença é a forma como são apresentados. O Plano A1 enfatiza a certeza de um bom resultado, ao passo que o Plano B1 liga a certeza a um resultado ruim. E, como demonstrou a pesquisa laureada com o Prêmio Nobel, feita por Daniel Kahneman e Amos Tversky, as pessoas confrontadas

com decisões preferem certezas, em lugar de apostas de risco sempre que um resultado seja bom – isto é, elas se tornam avessas ao risco, preferindo um bom desfecho garantido. No entanto, se o resultado oferecido parece ruim, as pessoas de repente passam a buscar o risco e alegremente se lançam no escuro para obter o resultado positivo. Qualquer pessoa que tenha consciência dessas características humanas pode cutucar os outros no sentido de tomar uma decisão específica simplesmente apresentando-a do jeito certo. Um consultor inescrupuloso, (imagine!) querendo que o cliente escolha um plano específico, deve enfatizar qualquer possível certeza de resultado positivo – em vez de se concentrar nas desvantagens certas, focalizar as vantagens incertas de outras alternativas.

Usar a teoria da decisão ajuda a nos vacinar contra essas estratégias, forçando-nos à dura e fria matemática. Como vimos, às vezes não há matemática a fazer: um conjunto de consequências domina o outro, independentemente do que de fato acontece. Vimos como essa dominância ajudou uma empresa a lidar com a ameaça potencial representada pelo uso de um produto químico supostamente arriscado. Mas ela também pode ser aplicada a questões bem mais importantes. Por exemplo, um dos maiores desafios que o mundo enfrenta hoje é como lidar com a ameaça do aquecimento global. Alguns argumentam em favor de medidas drásticas, como o abandono completo de combustíveis fósseis. Outros acham que devemos nos concentrar na adaptação ao clima em mudança, enquanto outros, ainda, insistem em que o aquecimento global é um mito – ou pelo menos que não tem nada a ver com ações do homem. Há bons motivos para acreditar que o aquecimento global está acontecendo, e com ele a mudança do clima em todo o planeta. Então, o que devemos fazer? Mais uma vez, a teoria da decisão nos ajuda a dissecar as complexidades para apresentar as opções de maneira incontroversa. Afinal, até o mais inveterado ambientalista ou cético da mudança climática pode ao menos concordar que o aquecimento global é uma realidade ou um mito. Construindo a usual matriz de decisão, vemos que todos os governos deveriam enfocar a redução do consumo de energia com a melhora na eficiência energética, pois essa é uma estratégia dominante – ou seja, faz

sentido, independentemente das realidades do aquecimento global (ver Tabela a seguir).

Essa conclusão é atualmente endossada por instituições como a Agência Internacional de Energia e a Fundação das Nações Unidas, que descrevem a conservação de energia como “o primeiro e melhor passo rumo ao combate contra o aquecimento global”. Todavia, durante décadas, ela parecia a Cinderela da estratégia energética global, ignorada pelos políticos. Talvez alguém deva lhes dar uma cartilha de teoria da decisão.

	O AQUECIMENTO GLOBAL É REAL	O AQUECIMENTO GLOBAL É UM MITO
DECISÃO A: CORTAR O CONSUMO DE ENERGIA PELO AUMENTO DA EFICIÊNCIA	<i>Consequência:</i> custos de implantação, mas substanciais impactos para retardar/impedir; para aperfeiçoamento; maior segurança energética.	<i>Consequência:</i> custos de implantação, mas conservação de recursos e dinheiro; e melhora na segurança energética.
DECISÃO B: NÃO FAZER NADA	<i>Consequência:</i> nenhum custo de implantação, mas grande impacto sobre saúde, economia, segurança global etc.	<i>Consequência:</i> nenhum custo de implantação, mas nenhuma reserva futura em recursos ou dinheiro nem melhora da segurança energética

Por que a conservação de energia é uma forma óbvia de combater o aquecimento global.

Conclusão

Muitas vezes as decisões envolvem uma traiçoeira mistura de probabilidades não claras e graves consequências. Anotar a gama de possibilidades e consequências numa tabela pode esclarecer o melhor curso de ação. Senão, sempre vale a pena tentar a aritmética básica da teoria da decisão.

18. Diga a verdade, doutor, quais as minhas chances?

QUANDO ALICE COMEÇOU a sentir dor no seio esquerdo, não quis correr riscos. Como mulher na casa dos sessenta anos, ela já vinha fazendo mamografias de dois em dois anos – e resolveu antecipar a seguinte, para descobrir a verdade o mais depressa possível.¹ Tirada a radiografia, ao deixar o centro médico, sentiu-se bem por ter feito a coisa certa, e a recepcionista lhe disse que ligariam para ela se houvesse algum problema. Alguns dias depois, o centro médico de fato ligou – e não para dar uma boa notícia a Alice. A mamografia tinha dado positiva. Ela ficou profundamente preocupada. Quem não ficaria? Uma rápida consulta na internet revela que as mamografias são precisas 80% das vezes. A implicação parece clara: há 80% de chance de que Alice esteja com câncer de mama. Isso é o que concluiriam muitos médicos.² Mas estariam errados – juntamente com o resultado positivo da mamografia –, porque isso só conta uma parte da história, a parte resumida, de maneira inadequada, pela noção aparentemente simples de “precisão”.

Para dar sentido a qualquer diagnóstico, a teoria da probabilidade revela que não precisamos de um, mas de *três* números. Dois deles refletem uma característica-chave de qualquer teste diagnóstico: seu potencial de induzir ao erro de duas formas distintas. Primeiro, ele pode detectar algo que na realidade não existe – produzindo o chamado falso positivo. Mas o teste também pode deixar passar algo que realmente existe, levando a um falso negativo. A capacidade de um teste evitar essas duas falhas é resumida por dois números: a taxa de verdadeiros positivos e a taxa de verdadeiros negativos – conhecidas tecnicamente (e com a típica opacidade) como *sensibilidade* e *especificidade*. Com os anos, têm-se tentado combinar as duas num número único, com a expectativa de representar a “precisão”, mas todas elas deixam a desejar de uma ou de outra maneira. Mantê-las separadas, por outro lado, permite-nos avaliar o

quanto devemos ficar impressionados com um diagnóstico. Afinal, qualquer médico pode diagnosticar, digamos, uma doença cardíaca de forma a abranger qualquer caso: simplesmente dizendo a todo paciente que ele tem um problema cardíaco. A taxa de verdadeiros positivos será de impressionantes 100%. Contudo, é óbvio que isso não é útil para diagnóstico – o que se reflete no fato de que a taxa de verdadeiros negativos (especificidade) é zero, porque o médico nunca diz a ninguém que a pessoa *não tem* problema cardíaco. O valor real do teste diagnóstico só pode ser avaliado conhecendo-se *as duas* taxas individualmente.

No caso da mamografia, as duas taxas são em torno de 80%. Isso quer dizer que, entre 100 mulheres com câncer de mama, a mamografia diagnosticaria corretamente a doença cerca de 80% das vezes, enquanto entre cada 100 mulheres sem a doença, mais ou menos 80% delas ouviriam que está tudo bem. Isso pode parecer confiável, mas, como acontece com tanta frequência quando se trata de probabilidade, a formulação verbal exata é problemática. O número de 80% de confiabilidade provém de testes em mulheres cuja condição de portadoras de câncer de mama já era conhecida. Como tal, ele mede apenas a confiabilidade do teste para confirmar o que já se sabia. Mas para mulheres como Alice, passando por exames de rotina, tudo o que sabemos de antemão sobre sua condição de câncer de mama provém de estimativas da *prevalência* da doença (ou “taxa de base”). Este é o terceiro número crucial de que precisamos para dar sentido ao resultado de um teste diagnóstico – e seu impacto pode ser drástico.

Mais uma vez, tomemos o caso de Alice. A prevalência de câncer de mama depende de uma legião de fatores, desde histórico étnico e perfil genético até idade, e, para dar sentido a qualquer resultado de teste individual, é vital utilizar o número apropriado. Por exemplo, o risco ao longo da vida para mulheres nos Estados Unidos é em torno de 12%, mas esse dado sofre desvio pelo enorme aumento do risco com a idade. Para mulheres com sessenta e poucos anos, como Alice, a prevalência é por volta de 5% – número que altera radicalmente as implicações de um resultado positivo da “mamografia 80% precisa”. Um pouco

de aritmética simples (ver Box a seguir) revela que, na verdade, há mais de 80% de probabilidade de o resultado positivo ser efetivamente um alarme falso. E é em grande parte o oposto exato das aparentes implicações de se obter um resultado positivo de um teste descrito como “80% preciso” – e mostra a importância de se levar em conta a plausibilidade de qualquer resultado de teste diagnóstico.

Como se deve reagir diante de um resultado positivo? Decerto faz sentido ficar um pouco preocupado: no caso de Alice, por exemplo, o resultado positivo do teste aumentava as chances de ela ter câncer de mama de 5% – a “taxa de base” para seu grupo etário – para 17%. Mas não há motivo para fatalismo ou pânico, pois até mesmo esse percentual mais alto significa que há 83% de probabilidade de não ser câncer de mama. A resposta apropriada é fazer outros testes, pois cada um acrescenta um peso de evidência a favor ou contra o diagnóstico de câncer de mama. E foi exatamente o que fez Alice – e, com toda a segurança, ela recebeu um ok.

Entretanto, nem sempre as coisas funcionam assim. Probabilidades não são certezas, e nunca se deve forçar demais a barra. Quando detectou um caroço no seio, a cantora Olivia Newton-John ainda estava com pouco mais de quarenta anos – e tinha um risco de câncer de mama que mal chegava a 1%. A mamografia deu negativa, bem como a biópsia. Mesmo assim, ela ia se sentindo cada vez pior, e afinal acabou descobrindo que tinha câncer. Menos de 1 em 10 000 mulheres da idade dela teria tido tanto azar a ponto de ter dois falsos negativos. Todavia, a teoria da probabilidade nos diz que, dadas oportunidades suficientes, mesmo eventos de baixa probabilidade se manifestam. Só que raramente ouvimos falar deles.

O mesmo raciocínio também mostra que toda mulher que passa por exames regulares deve se preparar para pelo menos um susto. O reverso da moeda dos 80% de confiabilidade de excluir aquelas que realmente não têm câncer é um risco de 20% de falsos positivos. No curso de mais ou menos dez testes bienais acima dos cinquenta anos, isso implica uma chance elevada de experimentar pelo menos um susto.

O QUE REALMENTE SIGNIFICA O RESULTADO DE UM TESTE “PRECISO”

Como técnica de diagnóstico para câncer de mama, a mamografia é bem impressionante: ela detecta cerca de 80% dos casos de câncer de mama, e dá ok para uma proporção daquelas mulheres livres da doença. Mas isso não nos diz precisamente nada sobre a probabilidade de Alice ter câncer, dado o resultado positivo de seu teste – porque não sabemos em qual desses dois campos ela se insere. Podemos, no entanto, ter alguma ideia a partir da prevalência do câncer de mama entre mulheres como ela. As estatísticas mostram que o risco para mulheres no seu grupo etário é em torno de 1 em 20. Então, vamos dar uma olhada nas implicações desses números brutos. Em 100 mulheres como Alice,

Número com câncer de mama: 5

Número sem: 95

Das cinco mulheres com câncer de mama, o verdadeiro positivo (“sensibilidade”) do teste detectará cerca de 80%, ou quatro mulheres. Mas o crucial é que elas não são as únicas a receber um resultado positivo. Das mulheres livres da doença, uma taxa de verdadeiro negativo (“especificidade”) de 80% significa que a maioria receberá corretamente um ok – mas ainda haverá 20% que não. Combinado com o fato de 95% não terem câncer de mama, isso leva a uma quantidade tenebrosa de falsos positivos:

Número de resultados positivos corretos: 80% de 5 = 4

Número de resultados positivos incorretos: 20% de 95 = 19

Então, o número total de resultados positivos é: 4 + 19 = 23

Podemos agora finalmente responder à pergunta-chave que Alice tinha a respeito do resultado positivo do teste: quais são as chances de que ela realmente tenha câncer?

$\text{Pr}(\text{câncer de mama, dado resultado positivo}) = \frac{\text{nº de positivos verdadeiros}}{\text{nº total de todos positivos}} = \frac{4}{23} = 17\%$

Então é de mais de 100 – 17 = 83% a probabilidade de que Alice esteja livre de câncer de mama, apesar da mamografia positiva.

Com uma quantidade cada vez maior de testes diagnósticos surgindo dos laboratórios de pesquisa, a necessidade de saber como interpretá-los nunca foi mais importante. No entanto, com muita frequência, até os pesquisadores

preferem deixar clara uma medida mais ou menos sem sentido de “precisão”, enquanto o papel das taxas de base é totalmente ignorado. Em julho de 2014, pesquisadores de duas universidades de ponta no Reino Unido divulgaram um exame de sangue supostamente “87% exato” em predizer o aparecimento da doença de Alzheimer entre pessoas com brandos problemas de memória. A história ganhou as manchetes na mídia e foi saudada como importante avanço pelo secretário de Saúde do governo britânico, Jeremy Hunt. Alguns pesquisadores, porém, sentiram a necessidade de contextualizar a história, e um especialista em Alzheimer advertiu que a impressionante “precisão” ainda significava que cerca de 1 em cada 10 pacientes teria o diagnóstico incorreto.

Na realidade, não é claro o que o número significava, pois os próprios pesquisadores nunca deixaram explícito o que entendiam por “precisão”. Dito isso, eles tiveram o grande cuidado de estabelecer os dois modos de falha do teste, conforme refletidos na sensibilidade e na especificidade. Usando dados coletados a partir de centenas de pacientes com várias formas de demência, descobriram que o exame de sangue predizia corretamente a progressão para o desenvolvimento pleno de Alzheimer em cerca de 85% dos casos, enquanto prognosticava corretamente a ausência de progressão cerca de 88% das vezes. Esses números implicam uma taxa de falsos negativos de 15% e de falsos positivos de 12%. Contudo, assim como nas mamografias, só podemos dar sentido a um resultado de teste positivo se avaliarmos sua plausibilidade – o que significa conhecer a taxa de base de risco de Alzheimer entre aqueles que fizeram o teste. Como este foi concebido para ser usado com pessoas que já tinham leve deficiência cognitiva, a taxa de base de risco é de cerca de 10-15%. Acionando a mesma aritmética que usamos para interpretar as mamografias, descobrimos que um exame de sangue positivo para Alzheimer implica uma chance de cerca de 50:50 de progressão para doença de Alzheimer. Então, como na mamografia, o número de “precisão” parece menos impressionante quando contextualizado. Os céticos podem alegar que o teste não é melhor que cara ou coroa, mas isso é injusto. Ao aumentar a probabilidade de progressão de 10-15% para 50%, o exame de sangue sem dúvida acrescentou peso de evidência

genuíno, o que a moeda nunca faz. Como tal, ele pode se tornar algum dia parte de uma bateria de testes para Alzheimer, como a mamografia e a biópsia são para o câncer de mama. Mas persiste o fato de que existe uma grande diferença entre a “precisão” do teste e as chances de Alzheimer implícitas num resultado positivo.

Os perigos de uma interpretação errada são extremamente agudos entre aqueles que decidem aplicar o teste em si mesmos usando kits de diagnóstico doméstico. Introduzidos pela primeira vez na década de 1970 para testar a gravidez, hoje é possível comprar kits de testes para muitas condições, de alergias a infecção pelo vírus da aids, o HIV. Como sempre, eles declaram-se impressionantemente “acurados”, mas o que isso significa e em que contexto, está longe de ser claro. No caso dos testes domésticos de gravidez, a precisão anunciada pode ser vista como algo bastante próximo do valor nominal: se der positivo, é altamente provável que você esteja grávida. Esses testes têm taxas baixíssimas de falsos positivos e falsos negativos – e, além disso, a maioria das mulheres que fazem os testes já tem fortes motivos para acreditar que estão grávidas. No entanto, até um teste de gravidez pode se mostrar altamente não confiável se feito por alguém que não deve estar grávido – como um homem, por exemplo. Em 2012, uma usuária do site de mídia social Reddit contou como um amigo homem tinha usado de brincadeira um teste de gravidez deixado no banheiro pela namorada – e ficou estarecido ao obter um resultado positivo.³ Como a taxa de base de gravidez entre homens é bastante baixa, não era provável que ele fosse dar à luz, apesar do resultado de um teste “preciso”. Mas esse não foi o fim da história. Outros usuários do Reddit contaram que o teste funciona detectando o hormônio HCG, produzido durante a gravidez da mulher – e nos casos de tumores testiculares. Uma visita ao médico confirmou o diagnóstico, levando a um tratamento precoce que deve ter salvado a vida do homem “grávido”.

A necessidade de considerar o fator plausibilidade num diagnóstico é importantíssimo com os kits para HIV. Eles também proclamam sua “precisão” bem superior a 90%. Entretanto, a menos que você tenha excelentes motivos

para acreditar que contraiu o vírus da aids, esse número é perigosamente enganador. Enquanto a especificidade e a sensibilidade são de fato superiores a 90%, a taxa de base de HIV fora dos conhecidos grupos de risco é muito baixa. Como consequência, resultados positivos para aqueles fora dos grupos de risco têm muito maior probabilidade de ser um alarme falso do que verdadeiros positivos.

Não é somente em diagnósticos médicos que o conceito de precisão deve ser considerado com cautela. O mesmo se aplica a qualquer teste alegando detectar sinais de algum traço de personalidade – como, por exemplo, ser mentiroso. Séculos atrás, na Ásia, acreditava-se que a desonestidade podia ser “diagnosticada” enchendo a boca dos suspeitos de arroz antes do interrogatório. Aqueles que tinham mais dificuldade em cuspir o arroz depois de questionados eram considerados culpados com base na alegação de que sua desonestidade lhes deixava a boca seca. Isso soa menos que confiável, e o que apreciamos antes a respeito dos diagnósticos médicos cristaliza essas dúvidas: ao mesmo tempo que o método pode ter uma taxa razoável de verdadeiros positivos, sua taxa de falsos positivos apresenta probabilidade de ser alta, dado que pessoas honestas também podem ficar com a boca seca por medo de não acreditarem nelas. Cabe pôr em contexto qualquer resultado positivo – o que exige uma estimativa das chances de que essa pessoa seja mentirosa, antes de se aplicar o teste.

Não que qualquer dessas coisas dissuada as pessoas que dizem ter inventado detectores de mentiras “precisos”. Desde a década de 1920, extrema atenção tem se concentrado nos chamados polígrafos, que monitoram uma porção de sinais fisiológicos, desde batimento cardíaco até sudorese, acreditando que revelem quando a pessoa está mentindo. Contudo, eles tiveram de superar o problema dos falsos positivos causados pelo estresse, ao mesmo tempo que foram enganados rotineiramente por espiões treinados. Aldrich Ames, o analista da CIA que trabalhou para os soviéticos durante as décadas de 1980 e 1990, passou repetidas vezes nos testes do detector de mentiras seguindo técnicas da KGB baseadas simplesmente em ter uma boa noite de sono, ficar calmo e ser simpático com os examinadores que operavam o polígrafo. Isso reduzia a taxa de verdadeiros

positivos, enquanto falsas pistas criadas pela KGB faziam ir por água abaixo a plausibilidade de Ames ser espião.

Entretanto, a busca para criar um detector de mentiras confiável não diminuiu. Em 2015, uma equipe conjunta de pesquisadores britânicos e holandeses anunciou uma técnica baseada na ideia de que pessoas culpadas ficam mais agitadas. Como sempre, reportagens na mídia focalizaram a impressionante “taxa de sucesso” de 82%, ao mesmo tempo sendo vagas em relação ao que aquilo significava. Um artigo superficial dos pesquisadores sugeria que ela era na realidade apenas a média da taxa de verdadeiros positivos, 89%, e da taxa de verdadeiros negativos, 75%. Se isso é verdade, trata-se efetivamente de um avanço significativo em relação ao polígrafo convencional, que, segundo os pesquisadores, tem um índice de acertos típico por volta de 60%. Todavia, como sempre, esses números deixam sem resposta a pergunta-chave: quais são as chances de alguém realmente ser mentiroso, dado um resultado positivo? Como sabemos agora, isso não pode ser respondido somente pelos resultados do teste: também precisamos ter algum conhecimento das chances de que o suspeito realmente seja um mentiroso, com base em evidências vindas de algum outro lugar. O que podemos dizer é que, se o “teste da agitação” for realmente confiável, como se alega, os que dão resultados positivos têm uma probabilidade maior de serem honestos do que de não serem, a menos que haja razões para pensar que as chances de serem culpados excedam aproximadamente a proporção de 1 em 5.

Exatamente o mesmo problema confronta tudo, desde testes de segurança em aeroportos e softwares de detecção de fraudes até alarmes contra invasões. Enquanto seus defensores se concentram na suposta “precisão” da tecnologia, essas alegações são desprovidas de sentido sem algum conhecimento da prevalência daquilo que está sendo buscado. E se essa prevalência for baixa – como, felizmente, é –, a única coisa que taxas incrivelmente altas de verdadeiros positivos e verdadeiros negativos previnem é uma enchente de alarmes falsos.

Há um modo simples de avaliar o resultado de testes para eventos raros, que chamamos de regra dos poucos por cento.⁴

A REGRA DOS “POUCOS POR CENTO”

Se você testar positivo para alguma coisa que afete menos do que poucos por cento das pessoas que fazem o teste, então é extremamente provável que o resultado seja um alarme falso, a menos que o teste seja tão bom que sua taxa de falsos positivos também esteja abaixo de poucos por cento.

Claro que todos iremos nos defrontar com falsos positivos durante nossa vida, desde exames médicos de rotina até revista de bagagem em aeroportos. Isso não é motivo para se tornar blasé. Boas decisões baseiam-se em combinar probabilidades e consequências; assim, uma pequena chance de uma consequência devastadora sempre deve ser levada a sério. Mas tampouco é motivo para uma reação exagerada. No final, nossa melhor proteção contra tanta coisa que tememos é sua mera improbabilidade. Como escreveu uma vez o industrial americano Andrew Carnegie: “Estou cercado de problemas a vida inteira, mas há uma coisa curiosa em relação a eles: nove décimos nunca aconteceram.”

Conclusão

Quando confrontado com o resultado de um teste diagnóstico, não se deixe enganar pela conversa acerca de sua “precisão”. Em muitos casos, o número que soa impressionante é apenas metade da história; muitos testes positivos têm maior probabilidade de se demonstrarem errados do que certos simplesmente pela raridade daquilo que tentam detectar.

19. Isso não é uma simulação!

Repito, isso não é uma simulação!

VIVENDO NUMA DAS regiões da Terra mais propensas a terremotos, os habitantes da Cidade do México são compreensivelmente ansiosos para ter um aviso precoce do próximo grande terremoto. Um dia, em julho de 2014, milhares deles receberam a notícia que temiam. Haviam baixado um aplicativo de celular que supostamente pegava dados da rede oficial de alertas sísmicos. Por volta da hora do almoço, o aplicativo enviou um alerta de que um grande terremoto estava prestes a acontecer. Em questão de segundos, as pessoas saíram correndo de seus locais de trabalho e ocuparam as ruas, protegendo-se da catástrofe. Esperaram, esperaram... e nada. Estava claro que havia sido um alarme falso. Os responsáveis pelo aplicativo emitiram um pedido de desculpas, dizendo que haviam interpretado mal uma mensagem da rede oficial. Então, mal se passaram dezoito horas, a cidade foi abalada por um forte terremoto de magnitude 6,3. O aplicativo de celular ficou silencioso.

De todas as calamidades naturais que os cientistas buscam prever, nenhum representa desafio maior que os terremotos. Até hoje não se encontrou nenhum meio de prever a hora, o local e a intensidade de um tremor de terra. E não é por falta de empenho. A busca de sinais denunciadores (“precursores”) de terremotos iminentes data de milênios. O escritor romano Claudio Eliano relata como os habitantes da antiga cidade grega de Helike notaram que ratos, cobras e muitas outras criaturas fugiam em massa cinco dias antes de um catastrófico terremoto destruir o local, no inverno de 373 a.C. Desde então, já se sugeriram inúmeros outros precursores, desde alterações no lençol freático até infiltrações de gases radiativos e mudanças em campos magnéticos. Alguns deles chegaram a ser levados a sério.

No inverno de 1975, a cidade de Haicheng, no nordeste da China, tornou-se cenário de estranhos eventos: o nível do lençol freático sofreu uma série de mudanças e as cobras emergiram de suas tocas. Depois a área foi atingida por um enxame de pequenos tremores. Acreditando que fossem abalos prévios pressagiando um terremoto muito maior, os geofísicos chineses emitiram o alerta de que um grande sismo estava prestes a ocorrer, e ordenou-se a evacuação em massa da cidade. Em 4 de fevereiro veio o terremoto: um devastador abalo de 7,3 na escala Richter. A medida fez com que quase toda a população da área atingida – estimada em 1 milhão de pessoas – sobrevivesse, com exceção de 2 mil vítimas. Parecia que o sonho da predição de um terremoto se tornara realidade. Isso até o ano seguinte, quando um tremor de 7,6 na escala Richter atingiu a cidade de Tangshan. Dessa vez não houve abalos prévios denunciadores, e pelo menos 255 mil pessoas morreram. Posteriormente alegou-se que houvera comportamento animal estranho na área. Seria essa a pista vital não observada, ou apenas uma racionalização posterior? Com que frequência os animais se comportam “de modo estranho” quando tudo está perfeitamente normal?

A resposta parecia óbvia: mais pesquisa. O enunciado da missão também parecia claro: se você não tem êxito, tente, tente de novo. Mas isso pressupõe que o êxito seja possível. E se não for? Na época do desastre de Tangshan, esse não era um ponto de vista popular. Muitos cientistas mantinham a fé de que seria possível montar uma rede capaz de detectar precursoros com horas e até semanas de antecedência, permitindo às pessoas ao menos se abrigar, se não fugir completamente da área. Evidente, os precursoros precisavam ser confiáveis, e os esforços para descobri-los foram redobrados. Mas uma questão atraía bem menos interesse: exatamente que nível de confiabilidade seria necessário – e haveria a probabilidade de qualquer precursor algum dia satisfazê-lo?

Enfrentar essa questão significava enxergar a predição de terremotos tal como ela realmente é: uma questão de diagnóstico confiável. Exatamente como os testes diagnósticos médicos se sustentam ou fracassam segundo sua capacidade de adicionar peso de evidência acerca de um risco específico, o

mesmo deve ocorrer com qualquer método de predição de terremotos. Se é para atender ao seu propósito, ele deve ser capaz de fazer a diferença entre alarmes falsos e o evento real. Mas, sobretudo, deve ser capaz de fazer isso de maneira a compensar o fato de que – felizmente – são raros os terremotos grandes o bastante para merecer evacuações em massa. A questão que então fica é: isso é remotamente plausível?

O resultado de se aplicar a “regra dos poucos por cento” do capítulo anterior não é encorajador, pois implica que, se o risco de um terremoto grande ocorrer ao longo de um período de, digamos, um mês for menor que poucos por cento – e certamente o é –, o alerta de terremoto provavelmente é um alarme falso, a menos que sua taxa de falsos positivos seja inferior a poucos por cento. Isso, por sua vez, exige que os precursores usados pelo sistema tenham essa taxa de falsos positivos. Apesar de séculos de esforço, jamais se encontrou algum precursor que sequer remotamente chegue perto desse padrão. Uma análise mais detalhada confirma essa desanimadora verdade.¹ Mesmo que, por algum milagre, se encontrasse algum precursor com uma taxa de 100% de verdadeiros positivos, a taxa de falsos positivos ainda precisaria ser menor que cerca de 1 em 1 000 para compensar a baixa probabilidade de ocorrência de um tremor grande. Nada do que foi descoberto sobre o processo pelo qual os terremotos são deflagrados tampouco dá qualquer esperança de se achar esse precursor confiável.

Pode-se perguntar por que um problema básico como a própria ideia de prever terremotos não eliminou de vez toda a busca décadas atrás. Os céticos mencionam as enormes verbas concedidas aos cientistas dispostos a procurar os obscuros precursores. Uma explicação mais caridosa é que os pesquisadores simplesmente não tinham consciência da barreira probabilística que coloca o êxito da procura para sempre fora de alcance. O assunto agora é trazido à baila, pois em meados dos anos 1990 a realidade começou a aparecer. O abjeto fracasso das tentativas de identificar precursores confiáveis tornou-se impossível de ignorar. Enquanto alguns persistem com o sonho de prever terremotos da mesma forma que os meteorologistas predizem tempestades, a maioria dos sismólogos desde então aderiu a um de dois campos.

O primeiro deles aceita que jamais será possível prever terremotos com a exatidão necessária, em termos de local e hora, permitindo que se adote uma atitude antes de qualquer evento específico. Em vez disso, ele concentra os esforços num fato incontroverso: algumas partes do mundo correm um risco inaceitável de serem atingidos pelos grandes terremotos. Não há dúvida, por exemplo, de que a maioria dos mais destrutivos abalos já registrados ocorreu em torno do chamado Anel de Fogo, que circunda o oceano Pacífico. Sabe-se também com segurança que algumas dessas áreas de alto risco se sobrepõem a zonas com alta densidade populacional, sobretudo o Japão. Logo, ao mesmo tempo que ninguém pode dizer exatamente quando e onde o próximo grande terremoto ocorrerá, sabemos os locais que enfrentam alto risco de ter grande número de vítimas. Os sismólogos do primeiro campo fizeram desses enunciados de alta confiabilidade a base para as chamadas estratégias de *mitigação* – construir prédios e instalações mais resistentes a terremotos, educar o público para responder quando acontecer o inevitável, o grande terremoto.

Tudo isso parece enfadonho perto do entusiasmo do tipo ficção científica que tem uma predição de tremor de terra, mas pelo menos funciona. Em fevereiro de 2010, o Chile foi atingido por um terremoto gigantesco, de 8,8 na escala Richter, um dos mais potentes já registrados. O evento causou estragos em todo o país e liberou energia suficiente para alterar a rotação da Terra. Contudo, menos de 600 pessoas morreram – em grande parte como resultado dos códigos de edificação do país, exigindo que a resistência a terremotos seja um fator incorporado às construções domésticas e comerciais. Em contraste agudo e trágico, o país caribenho do Haiti foi atingido por um terremoto bem mais fraco, de 7 na escala Richter, algumas semanas antes. Apesar de ser 500 vezes mais fraco, o abalo desmoronou cidades cheias de barracos amontoados e malconstruídos, matando 220 mil pessoas.

Em essência, a estratégia de mitigação tem sucesso por se concentrar nas escalas temporais em que os terremotos são basicamente uma certeza – eliminando assim a necessidade de precursores de confiabilidade impossível. Há, porém, outra estratégia que também tem se mostrado muito bem-sucedida.

Ironicamente, ela utiliza o precursor definitivo de qualquer terremoto: o próprio abalo sísmico.

Os terremotos começam quando a rocha não aguenta mais o esforço a que está submetida e se rompe. O ponto em que a ruptura começa é conhecido como foco, e é daí que as ondas sísmicas se alastram, provocando a destruição. Essas ondas, no entanto, vêm em diferentes formas – e, o que é mais importante, não viajam com a mesma velocidade. As mais rápidas são as chamadas ondas P ou primárias, movimentos de vaivém que viajam numa velocidade incrível de cerca de 10 mil a 20 mil quilômetros por hora. Depois vêm as ondas S ou secundárias, movimentos de vaivém muito mais destrutivos – contudo, viajam apenas com a metade da velocidade. Logo, ao se detectarem ondas P, é possível enviar um alerta de terremoto extremamente confiável entre 30 e 60 segundos antes de ele chegar. Isso pode não parecer muito, porém é o suficiente para salvar vidas, como reconheceram os engenheiros japoneses na década de 1960, quando construíram a famosa rede Shinkansen do “trem-bala”. Eles instalaram sismógrafos que alertam os maquinistas dos trens para acionar os freios e reduzir o risco de descarrilamentos em alta velocidade. No começo dos anos 1990, isso havia se transformado no Sistema Urgente de Detecção e Alarme de Terremotos (UrEDAS), que identifica ondas P e automaticamente assume o controle dos trens em perigo. O sistema não é infalível: em 2004 um trem-bala ao norte de Tóquio descarrilou depois de ser atingido por ondas S de um terremoto de 6,8 Richter, cujo epicentro estava perto demais para se fazer alguma coisa. Mesmo assim, o sistema confirma a necessidade de “precursores” extremamente confiáveis, mesmo que só possam dar avisos com poucos segundos de antecedência. Assombrosamente, não há uma única morte relacionada aos abalos na rede do trem-bala em mais de cinquenta anos de operação e viagens de 10 bilhões de passageiros numa das partes do planeta de maior atividade sísmica.

O sistema UrEDAS agora foi ampliado para todo o país, a fim de pelo menos emitir o alerta alguns momentos antes. Os que estão dentro de casa podem se proteger afastando-se das paredes externas e das janelas e ficando debaixo de mesas; os que estão na rua podem tentar chegar a espaços amplos e abertos.

Durante o devastador terremoto de 9,0 Richter que atingiu Fukushima em março de 2011, uma rede de televisão enviou alertas pressagiando a chegada de ondas de choque com um minuto ou mais de antecedência, e salvou muitas vidas. Sistemas de alarme semelhantes estão difundidos em outros lugares, especialmente no México. Uma rede de detectores sísmicos foi montada ao longo da costa de Guerrero, a 350 quilômetros da Cidade do México, e dá alertas com antecedência de cerca de um minuto. Combinados com a mitigação dos efeitos de terremotos, esses sistemas de alerta agora têm êxito onde o sonho da ficção científica fracassou.

Os mesmos conceitos que provocaram a morte desse sonho deram lições para a predição do mais inconstante dos fenômenos naturais, o clima. Aqui sem dúvida se fez progresso. Segundo a Agência Meteorológica do Reino Unido, progressos em monitoramento de satélite e computação levaram as previsões do tempo de quatro dias a se tornar tão confiáveis quanto as previsões de um dia na década de 1980. Números com precisão de 70 a 80% são considerados para previsões de sol e chuva, e com mais de 90% para amplitudes de temperatura. Como sempre, não está claro aqui o que se entende por “precisão”. Em todo caso, muitos britânicos se empenhariam para comparar os números com sua experiência de serem apanhados de surpresa em aguaceiros imprevisíveis ou preparar-se para tempestades que simplesmente não acontecem.

O problema aqui é menos a “não confiabilidade” das previsões de tempo e mais o nosso fracasso em saber como reagir a elas. Imagine, por exemplo, que você esteja planejando passar sua hora de almoço no parque, e ouve que a previsão é de chuva. Sabendo que as previsões de chuva são cerca de 80% acuradas, parece óbvio que você deve pelo menos levar o guarda-chuva. Todavia, isso ignora o fato de que a precisão vem de duas formas: predizer corretamente algo que é verdade e ignorar corretamente algo que é falso. No caso das previsões de chuva, vamos supor que a proporção de 80% valha tanto para verdadeiros positivos quanto para verdadeiros negativos. Então, sabemos que em 100 casos nos quais efetivamente chove, a previsão estará certa em 80 casos; e em 100 casos em que o tempo fica seco, a previsão acerta 80 vezes. Para

saber como reagir à previsão, porém, ainda precisamos de mais um número: as chances de que chova durante o nosso intervalo de almoço. Para o Reino Unido, a probabilidade de chuva a qualquer hora é de aproximadamente 10%. Agora temos tudo que necessitamos para calcular o que realmente significa a previsão. E não é absolutamente o que se esperava.

O jeito mais fácil de ver isso é calcular o que aconteceria em 100 casos da nossa situação: uma hora passada ao ar livre quando a previsão é de chuva. Sabemos que desses 100 casos, a chuva normalmente é esperada em mais ou menos 10, e o tempo permanece seco nos outros 90. A taxa de 80% de verdadeiros positivos da previsão significa que, dos 10 dias de almoço molhado, a meteorologia vai prever corretamente chuva em 8 deles. Mas não será a única vez que haverá previsão de chuva. A taxa de 80% de verdadeiros negativos significa que haverá previsão incorreta de chuva em 20% dos casos em que não chove. Isso é 20% dos 90 almoços secos, implicando que a meteorologia preverá chuva incorretamente em mais 18 ocasiões. Então, no total serão $8 + 18 = 26$ predições de chuva, das quais apenas 8 serão verdadeiros positivos – uma taxa de ocorrência de $8/26$ ou 31%. Isso é muito abaixo do que esperaríamos se simplesmente tomássemos ao pé da letra a alegação do Serviço de Meteorologia de 80% de acurácia. Mas mostra a crucial importância de incluir o fator plausibilidade de qualquer previsão – nesse caso, determinado pelos baixos 10% de risco de chuva.

Entretanto, ainda nos resta uma decisão: saímos para dar o nosso passeio, levamos guarda-chuva ou esquecemos a coisa toda? O senso comum sugere que a resolução depende das consequências de se ignorar a previsão, mas a coisa não para aí. Assim como as previsões podem ser inexatas de duas formas diferentes, nossa reação a uma previsão também pode ser errada de duas maneiras diferentes. Quanto ao clima, por exemplo, podemos ignorar uma previsão que se revela correta ou confiar numa previsão que se mostra errada. A melhor decisão acaba dependendo de uma interação surpreendentemente complexa entre a prevalência do evento em si, a confiabilidade da previsão e a nossa visão sobre as consequências de tomar a decisão errada. Em outras palavras, o que seria a

reação certa a uma previsão “precisa” para uma pessoa pode ser errada para outra. Por exemplo, acionando a matemática,² os números que vimos implicam que você deveria ignorar a previsão de chuva – a menos que você ache que tomar chuva é *pelo menos* duas vezes pior que a frustração de cancelar seu passeio e acabar descobrindo que afinal não choveu. E a ideia de levar o guarda-chuva? Você não deve se incomodar com isso, a não ser que considere ser surpreendido sem guarda-chuva *pelo menos* duas vezes mais incômodo que carregá-lo e descobrir que não tinha necessidade dele.

Depois disso tudo, não é de admirar que as previsões tenham péssima reputação. Até quando há precursores bastante confiáveis, a previsão em si pode se mostrar pior que inútil – simplesmente porque está tentando predizer algo (em geral, felizmente) incomum. A maioria de nós também tem uma compreensão errônea do conceito de “precisão”, e fazemos escolhas à luz da predição. Mesmo assim, nos sentimos livres para culpar os responsáveis pela previsão quando as coisas não saem conforme o esperado. E pairando acima de tudo está o fato mais básico de todos: estamos lidando com incerteza e probabilidade. Assim, os benefícios de confiar em métodos de previsão comprovados só surgem com o tempo, e não o tempo todo.

Conclusão

O sonho de prever eventos naturais é tão antigo quanto a história, mas nossa capacidade de realizá-lo é restringida por limites fundamentais. Saber quais são, e como o método de previsão lida com eles, é a chave para tomar decisões ideais acerca de eventos futuros.

20. A fórmula milagrosa do reverendo Bayes

QUANDO A Guarda Costeira dos Estados Unidos foi alertada do que tinha acontecido na costa de Long Island numa noite de julho de 2013, parecia evidente que a missão de busca e salvamento teria um desfecho trágico. Uma lagosteira reportara que um membro de sua tripulação sumira do barco a cerca de 60 quilômetros em alto-mar, no Atlântico.¹ De algum modo ele tinha caído da embarcação durante a noite. Pior ainda, estava trabalhando sozinho, e ninguém sabia exatamente quando ou onde ocorrera o acidente. Quando o piloto do helicóptero, o tenente Mike Deal, e seus colegas decolaram, sabiam que as chances de localizar uma pessoa flutuando em algum ponto numa área de mais de 4 mil quilômetros quadrados de oceano eram muito pequenas. Mas não eram nulas – e isso dava esperança à equipe, por causa do incrível kit com o poder de ampliar drasticamente as oportunidades de sucesso conhecido como Sarops (de Search and Rescue Optimal Planning System). Isso pode parecer uma sofisticada caixa repleta de sensores, componentes eletrônicos e microchips, mas o Sarops é na verdade um algoritmo: uma receita matemática capaz de processar até pistas muito vagas sobre quando e onde um marinheiro teve problemas e combiná-las com conhecimentos sobre as condições locais, de modo a restringir radicalmente a área de busca.

Naquela manhã de julho, a Guarda Costeira alimentou o Sarops com estimativas da hora provável em que o pescador caiu do barco, e o dispositivo respondeu com os lugares mais indicados onde procurar. Armados dessas informações, o tenente Deal e seus colegas entraram no helicóptero e começaram a busca. À medida que passavam as horas, novas informações surgiam a respeito da hora provável do acidente, e o Sarops produzia mapas atualizados para a equipe do helicóptero. Finalmente, depois de sete horas e com o marcador de combustível dizendo para retornar à base, o copiloto do tenente Deal soltou uma

exclamação. Havia localizado algo. Eles deram meia-volta – e lá estava o pescador no meio da ondulação do oceano, acenando freneticamente.

Dadas as chances de sucesso, o que o Sarops permitiu ao tenente Deal e seus colegas naquele dia quase parece um milagre. E de fato ele lança mão de ideias exploradas por um clérigo. Não está claro o que exatamente estimulou o ministro presbiteriano inglês e matemático amador Thomas Bayes (1702-1761) a desenvolver a fórmula que leva seu nome. Mas não existe qualquer dúvida de que o resultado veio a se tornar um dos mais controversos na teoria da probabilidade, cuja simplicidade e apelo intuitivo escondem um assombroso poder.² Uma pista para todo o alvoroço pode ser encontrada num enunciado simples do que está implícito no trabalho de Bayes:

A REGRA MILAGROSA DO REVERENDO BAYES

Novo nível de crença sobre algo = velho nível de crença + peso da nova evidência

Para algo supostamente baseado nas leis da probabilidade, este é um enunciado muito esquisito, e nele não há nada sobre probabilidades, frequências ou aleatoriedade. Em vez disso, diz respeito aos mais intuitivos conceitos de crença e evidência. E ressalta uma característica da probabilidade que Bayes reconheceu, mas que continua controversa até hoje: a de que ela pode ser usada para captar graus de crença. Até agora, nós focalizamos quase inteiramente a probabilidade em seu papel familiar de compreender eventos do acaso, como jogar dados. Mas, como vimos no capítulo sobre os cassinos, essa é na verdade apenas uma das formas da fera: a probabilidade “aleatória” (que vem de “jogador de dados”, em latim). O trabalho de Bayes revelou um uso muito mais potente do conceito, como meio de captar a incerteza causada não pela aleatoriedade, mas pela simples falta de conhecimento. Essa incerteza “epistêmica” (da palavra grega para “conhecimento”) é muitíssimo diferente porque, pelo menos em princípio, podemos reduzi-la usando evidências.

A questão de como e de quanto é o foco do trabalho de Bayes. Como tal, tem um peso direto na busca que está no coração de todo empreendimento científico: transformar evidência em conhecimento confiável. Não que se pudesse adivinhar isso a partir do título da obra de Bayes sobre o tema. Publicado em 1764, *Essay Towards Solving A Problem in the Doctrine of Chances* soa maçante. Escrito em inglês arcaico e recheado de álgebra antiquada, é difícil para os olhos, quanto mais para o cérebro.³ Acrescente-se o fato de que o próprio Bayes jamais conseguiu publicá-lo, e é outro milagre sabermos alguma coisa sobre ele. Por isso, temos uma dívida para com o amigo de Bayes e seu colega, o matemático amador Richard Price, que o encontrou em meio à papelada do reverendo logo após a morte deste, em 1761. Reconhecendo as implicações do ensaio, Price chamou a atenção da Royal Society, a mais importante academia científica do mundo. A Royal Society o publicou devidamente, junto com uma introdução escrita por Price, que estava determinado a assegurar que a importância do texto fosse reconhecida. Ele ressaltou que Bayes havia atacado um problema que “de maneira nenhuma era apenas uma especulação curiosa na doutrina das probabilidades”, mas tinha uma relevância direta em “todos os nossos raciocínios concernentes a fatos passados e o que se torna provável daí em diante”.

Não se sabe exatamente por que Bayes nunca publicou o ensaio. Talvez tivesse sentido que havia mais trabalho a fazer, mas carecia de poder de fogo matemático para realizá-lo. É improvável que adivinhasse como suas anotações levariam a uma controvérsia ainda em ebulição 250 anos depois, com alguns pesquisadores evitando usar o termo “bayesiano” em seus artigos acadêmicos por medo de provocar brigas.

Leitores astutos podem desconfiar de que a fonte de encrenca reside nos ingredientes exigidos pela regra de Bayes para atualizar as crenças. Veremos em breve a verdade disso, mas entender sua origem pode ser útil. Eles provêm da tentativa de Bayes de responder a uma pergunta perfeitamente razoável, que não foi dada pelos brilhantes matemáticos que tinham fundado a teoria da probabilidade no fim do século XVII. Eles haviam concebido fórmulas que

davam a probabilidade de vários eventos aleatórios clássicos – por exemplo, de tirar três 3 seguidos em dez lançamentos de dado. Essas fórmulas tinham valor óbvio para os jogadores, que podiam usá-las para decidir se valia a pena aceitar uma aposta. Tudo que tinham a fazer era alimentar a fórmula⁴ com três números: as chances de obter o evento em questão em um lançamento qualquer (nesse caso, $1/6$), o número de sucessos a ser obtido (3) no número total de tentativas (10), e a resposta simplesmente surgiria (algo em torno de 1 vez em 6,5). Se o jogador achasse alguém disposto a oferecer condições *menos* prováveis que essa de o evento ocorrer – digamos, de 1 em 10 –, a aposta valia a pena, pois significava que a pessoa que oferecia essas condições achava o evento menos provável do que ele realmente é – possivelmente por ignorar os cálculos. O jogador, porém, devia ser cauteloso, pois a mesma fórmula podia ser usada por agentes de apostas espertos a fim de oferecer condições para apostas enganadoras similares, com chances radicalmente diferentes – como as de obter *pelo menos* três 3 (cerca de 1 em 4,5) ou *não mais que* três 3 – que, com uma chance de 93%, é praticamente certa. A fórmula podia lidar com tudo isso; ao mesmo tempo, cabe ressaltar o fato de que a formulação correta é importante em probabilidade – o que, como veremos, tem gerado imensa controvérsia concernente ao trabalho de Bayes.

Em face disso, o objetivo de Bayes era perfeitamente claro e direto. Ele quis pegar as fórmulas usuais e invertê-las. Isto é, em vez de começar com um conhecimento do que, digamos, um dado pode fazer e depois calcular as chances de diferentes resultados, Bayes queria começar com os resultados, e então trabalhar de trás para a frente a fim de ver o que eles revelavam acerca do dado. Está claro que a fórmula para isso também seria útil para jogadores – no mínimo para detectar trapaças. Depois de ver alguém tirar quatro 6 em cinco tentativas, poderíamos desconfiar de alguma trapaça, mas como poderíamos usar a evidência para quantificar nossas desconfianças?

No seu *Ensaio*, Bayes apresenta a teoria para fazer esse cálculo. Ele começa provando uma receitinha elegante para lidar com uma pergunta muito comum: como calculamos a probabilidade de eventos cuja aparição é influenciada por

eventos anteriores? Por exemplo, se acabamos de tirar um ás de um baralho e não o colocamos de volta, isso claramente afeta as chances de tirar outro ás. Bayes deduziu a fórmula necessária (ver Box).

COMO O TEOREMA DE BAYES TRANSFORMA INFORMAÇÃO EM CONHECIMENTO

A forma mais básica do teorema de Bayes mostra como as chances de um evento A ocorrer afetam as chances de um evento subsequente B. Especificamente, a “probabilidade condicional” de B, dado A, é:

$$\Pr(B | A) = \Pr(A | B) \times \Pr(B)/\Pr(A)$$

Isso permite que informação nova se torne conhecimento. Por exemplo, se alguém tirar ao acaso uma carta do baralho, nós sabemos sem olhar que as chances de ser de ouros são de 1 em 4. Mas se nos disserem que a carta é vermelha, o teorema de Bayes mostra que as chances de ser de ouros saltam para $\frac{1}{2}$. Isso porque $\Pr(\text{vermelho} | \text{ouros}) = 1$ (pois todos os ouros são vermelhos), $\Pr(\text{ouros}) = \frac{1}{4}$ e $\Pr(\text{vermelho}) = \frac{1}{2}$, então pelo teorema de Bayes, $\Pr(\text{ouros} | \text{vermelho}) = \frac{1}{2}$. Claro que não precisamos realmente do teorema de Bayes para fazer esse raciocínio, pois todo mundo sabe que metade das cartas do baralho são vermelhas. A questão é que a mesma ideia básica funciona com problemas bem mais complexos – tais como diagnósticos médicos.

Mas há outro ponto simples, mas importante, digno de nota: o perigo de distraidamente trocar as probabilidades condicionais entre si: $\Pr(B | A)$ pode parecer semelhante a $\Pr(A | B)$, mas o teorema mostra que só serão idênticas se $\Pr(B)$ também for igual a $\Pr(A)$. Como veremos, isso é crucial para compreender um grande escândalo que manchou a ciência durante décadas.

Bayes foi adiante para mostrar como essa fórmula simples podia oferecer um modo de transformar observações em conhecimento. Por exemplo, qualquer um que presencie a moeda numa proporção inusitadamente alta de lançamentos pode usar a fórmula e transformar essas observações em conhecimento acerca da honestidade da moeda, em específico as chances de o resultado esperado ser realmente em torno de 50%. Mas, como ressalta Price na introdução ao livro do amigo falecido, Bayes assentara o alicerce para muito mais do que sugeria o enfadonho título: ele abriu caminho para atacar o problema *geral* de transformar

observações em conhecimento. O elo vem da maneira como as probabilidades podem ser usadas para avaliar níveis de crença. Rotineiramente, fazemos a ligação em conversas do dia a dia: falamos em acreditar que algo tem “grande chance” de acontecer, de lidar com uma “chance de 50:50”, de ter “99% de certeza” sobre um fato. O que Bayes fez foi mostrar que não só é possível quantificar nossas crenças como probabilidades ou chances, sua prima próxima, como também podemos também aplicar a elas as leis da probabilidade.

Embora ele nunca o tenha enunciado dessa maneira, o teorema que leva seu nome pode ser reescrito de uma forma que nos permita atualizar nossas crenças à luz de nova evidência (ver Box a seguir).⁵

Em termos simples, o teorema de Bayes mostra que é possível captar nosso nível de crença em alguma hipótese ou argumento usando a linguagem da probabilidade. O teorema assume sua forma mais simples quando são usadas *chances de erro* para exprimir algum grau de crença de que uma teoria se mostre correta ou não. Teorias plausíveis – como a ideia de que o sol vai nascer amanhã – têm sua plausibilidade captada por probabilidades elevadas, e, portanto, chances de erro “pequenas”, enquanto alegações implausíveis (digamos, de Elvis habitar o lado escuro da Lua) têm baixa probabilidade, e, portanto, chances de erro grandes. O teorema de Bayes mostra que podemos atualizar nosso nível de crença inicial (“a priori”) à luz de nova evidência multiplicando-o por um fator conhecido como razão de probabilidade (RP). Esta última expressa o peso da evidência fornecida por, digamos, um experimento de laboratório ou um estudo de longo prazo com muitas pessoas. A RP pode parecer complexa, mas também é intuitiva. Por exemplo, se a probabilidade de obter a evidência que vimos, *dado* que a nossa crença esteja correta, é muito alta, o numerador (número acima do traço de fração) será próximo de 1, o valor mais alto que qualquer probabilidade pode assumir.

TEOREMA DE BAYES: COMO ATUALIZAR SUA CRENÇA COM EVIDÊNCIAS

O teorema de Bayes mostra como as chances de sua crença num argumento ou teoria específica devem mudar à luz de nova evidência. A forma mais simples do teorema mede o impacto da evidência sobre as *chances* de o argumento se provar correto:

$$\text{Chances (sua crença estar correta, dada a nova evidência)} = \text{RP} \times \text{Chances (sua crença estar correta),}$$

onde RP é a “razão de probabilidade”. É ela que capta a força da evidência que você descobriu, e é determinada tomando-se a razão de duas “probabilidades condicionais” – isto é, probabilidades que dependem de duas premissas concorrentes.

$$\text{RP} = \frac{\text{Pr (a evidência ser observada, presumindo que sua crença seja correta)}}{\text{Pr (a evidência ser observada, presumindo que sua crença NÃO seja correta)}}$$

Embora isso pareça complicado, na verdade é bastante intuitivo e fácil de usar, uma vez explicado honestamente (ver o texto principal e os exemplos).

A RP reflete o fato de que a evidência consistente com nossa crença carrega mais peso que a evidência irrelevante para nossa crença, ou mesmo contrária a ela (como é refletido por uma probabilidade de menos de 0,5). Se, além disso, existe apenas uma chance *baixa* de obter a evidência que vimos, dado que nossa crença está *errada*, isso significa que a evidência faz o bom serviço de discriminar entre nossa crença e outras possibilidades. Mais uma vez, como sugere o senso comum, isso aumenta o peso da evidência respaldando nossa crença por meio de um aumento total do valor de RP.

Para dar um exemplo, se nossa crença de que, digamos, uma paciente tem câncer de mama é de apenas 5% antes de chegarem os resultados dos exames, fixaríamos o nível de crença inicial como proporção de câncer de 0,05. Suponha, então, que o exame seja feito por um método para o qual a probabilidade de obter resultado positivo, admitindo que haja câncer de mama, seja de 80%, enquanto a chance de obter positivo, admitindo que não haja câncer de mama (a

“taxa de alarme falso”), seja de 20%. Sabemos assim que o método de exame tem uma razão de probabilidade de $0,8/0,2 = 4$, e o teorema de Bayes nos diz que um resultado de exame positivo deve aumentar nosso nível de crença inicial de que a paciente tenha câncer de mama em 4 vezes as proporções iniciais de 0,05, dando uma proporção atualizada de 0,2. Traduzindo de volta para a probabilidade, essa proporção implica uma chance de 17% – logo, uma chance de 83% de a paciente ainda estar livre da doença, apesar do resultado positivo. Esse resultado chama nossa atenção porque é exatamente o que obtivemos usando proporções simples e senso comum, no Capítulo 18. E isso esclarece um fato-chave sobre Bayes: em situações nas quais toda a informação necessária é bem definida e mensurável, e a gama de resultados possíveis é bastante simples (como câncer/não câncer), não há nada de remotamente controverso em relação ao seu teorema.

Mas como o próprio Bayes reconheceu, em muitos usos potenciais do seu prático teorema as coisas não são tão claras e diretas. Os leitores astutos já podem ter percebido por que a regra de Bayes mostra como atualizar crenças à luz de nova evidência. Mas isso requer, em primeiro lugar, alguma crença a priori a ser atualizada. No caso do exame de câncer, podemos obter nosso nível a priori de crença sobre as chances de alguém ter câncer de mama a partir de estudos amplos da população como um todo. A Guarda Costeira dos Estados Unidos que procurava o pescador perdido claramente se defrontou com um problema mais traiçoeiro. Primeiro, a crença não era uma simples dicotomia tipo verdadeiro/falso; era sobre em que áreas seria melhor procurar. Depois, havia o problema de que as equipes de busca só tinham alguma crença prévia vaga de onde o acidente teria acontecido. Mas foram capazes de tirar proveito da característica-chave do teorema de Bayes: sua capacidade de atualizar continuamente as crenças. Logo, quando os palpites iniciais sobre o paradeiro do pescador se mostraram incorretos, as crenças – mais as novas informações sobre correntes e fatores similares – puderam ser alimentadas de modo a se tornar o novo nível inicial de crença, que foi por sua vez atualizado numa série de iterações que acabaram acertando em cheio no alvo.

Todavia, pelo menos a Guarda Costeira tinha alguma ideia vaga de onde começar. Claro que não havia sentido em procurar, digamos, no Pacífico. Mas o que fazer se não houver nenhuma boa evidência? Imagine que estejamos participando de algum jogo novo no cassino, e desconfiamos de que ele talvez esteja viciado. Como captar nossa crença inicial (“a priori”) de que o jogo é desonesto, dado que não temos realmente nada em que nos basear? O próprio Bayes sugeriu uma maneira de lidar com esse assunto – denominado, de forma pouquíssimo imaginativa, de “problema dos a priori” – usando as observações disponíveis para dar o primeiro palpite. A coisa funcionava, mas só em certas circunstâncias. Isso fez com que julgassem sua fórmula apenas com um valor restrito,⁶ e nem os esforços de Price para promover o trabalho do amigo e sua publicação pela mais famosa instituição científica conseguiram impedir que ele fosse quase inteiramente ignorado.

Felizmente – como acontece tantas vezes com as descobertas importantes – Bayes não estava sozinho ao pensar sobre como transformar dados observacionais em conhecimento. Como um dos mais brilhantes expoentes da aplicação da matemática aos problemas da vida real, Pierre Simon de Laplace vinha refletindo sobre as mesmas questões durante anos quando, em 1781, soube do trabalho de Bayes por meio de um colega. Ele também vinha lutando com o “problema dos a priori”, e deparou com uma solução aparentemente óbvia: se não temos conhecimento nenhum sobre, digamos, as chances de uma moeda específica dar cara, por que não admitir simplesmente que ela tem igual probabilidade de assumir um valor entre 0 e 100%? Conhecido como “princípio da razão insuficiente” ou “princípio da indiferença”, ele era simples de utilizar e parecia estar aberto a um sem-número de aplicações.

O próprio Laplace se propôs a usar a fórmula para atacar problemas em tudo, desde demografia e medicina até astronomia. Na época da sua morte, em 1827, ele tinha dado à regra de Bayes o formato moderno e o *imprimatur* da autoridade (de fato, pode-se argumentar que a proposição devia se chamar teorema de Bayes-Laplace). Mas logo seus métodos se viram sob o ataque de uma nova geração de pesquisadores, que se concentraram naquilo que consideravam o

calcanhar de aquiles de todo o processo: o problema de estabelecer níveis a priori de crença na ausência de qualquer evidência. Alguns eram contrários ao uso da “indiferença” de Laplace como ponto de partida para os cálculos; outros não gostavam do emprego de probabilidades como vagos “níveis de crença”, em vez de belas e concretas frequências de eventos.

A crítica mais contundente veio daqueles que viam o teorema de Bayes-Laplace como ameaça a todo o empreendimento científico. Para eles, a necessidade que o teorema impõe de um enunciado de crença inicial ameaçava o aspecto mais adorado da pesquisa científica, sua objetividade. Na ausência de qualquer conhecimento a priori, o que impediria os pesquisadores de pegar dados observacionais e tirar qualquer conclusão, simplesmente ajustando o nível a priori de crença para obter o resultado desejado? Que cientista com respeito próprio poderia assistir a isso calado e permitir que tais práticas abrissem caminho para penetrar na desapaixorada busca da verdade?

Nos anos 1920, o teorema de Bayes havia sido excomungado da pesquisa científica. Mesmo que os estatísticos mais influentes da época aceitassem a pequena e elegante receita de Bayes para calcular “probabilidades condicionais” de eventos afetados por outros eventos, eles rejeitavam sua função de transformar evidência em conhecimento. Em vez disso, eles conceberam toda uma caixa de ferramentas de conceitos “frequentistas” aparentemente objetivos, em que as probabilidades eram apenas as frequências com que os resultados ocorriam em dada oportunidade. Em resumo, eles tentavam evitar o problema dos a priori suscitado pelo teorema de Bayes atendo-se às fórmulas originais da teoria da probabilidade, que simplesmente forneciam os resultados esperáveis, *presumindo* que a causa já fosse conhecida. Por exemplo, os frequentistas se propunham a investigar se uma moeda era honesta *presumindo* que ela o fosse, e usavam as fórmulas da probabilidade para calcular o que devia se observar *caso* a premissa fosse verdadeira. Se aquilo que era observado tivesse apenas uma chance muito pequena de acontecer caso se tratasse de uma moeda honesta, os frequentistas argumentavam que isso evidenciava que a chance de a moeda ser honesta também era muito pequena, portanto, devia-se desconfiar de trapaça.

Se para você isso não soa muito certo, parabéns. Você acabou de identificar uma falha de raciocínio que muitos pesquisadores – talvez a maioria – deixaram de perceber no último século e tanto. O argumento comete a fundamental tolice de alegar que a probabilidade de A, dado B, é a mesma que a de B, dado A. No exemplo citado, o erro específico está em assumir que é correto argumentar que:

$$\text{Pr}(\text{evidência de lançamentos, dada moeda honesta}) = \text{Pr}(\text{moeda honesta, dada evidência de lançamentos})$$

Mas, como Bayes mostrou com os resultados absolutamente incontroversos referentes às probabilidades condicionais, trocar as proposições de lugar é uma coisa muito perigosa. Como vimos no exemplo de como atualizar as crenças a partir da evidência, com questões simples de probabilidade, isso leva a resultados simples e totalmente errados – como a probabilidade de que uma carta seja de ouros, sabendo-se que ela é vermelha (50% de probabilidade), ser igual à probabilidade de que uma carta seja vermelha, sabendo que é de ouros (cuja probabilidade é 100%). Entretanto, quando usada para transformar evidência em conhecimento, a descuidada troca de lugar das probabilidades condicionais é uma receita de desastre, porque comete a falácia lógica de primeiro presumir que algo seja verdade para chegar a uma dedução, e depois usar a dedução para testar a premissa.

O teorema de Bayes mostra que a única maneira de trocar as probabilidades condicionais de lugar é introduzindo informação adicional. No caso de tirar inferências sobre nossas crenças a partir de dados, isso representa incluir alguma probabilidade a priori de que a nossa crença esteja correta, o que, por sua vez, significa que precisamos enfrentar o problema dos a priori. Como vimos, esta nem sempre é uma questão: às vezes há uma fonte óbvia para o conhecimento a priori – como na pesquisa mencionada. Contudo, muitas vezes não há fonte, e devemos encarar o fato de que tirar conclusões a partir de dados talvez envolva um trabalho subjetivo de adivinhação. No entanto, o teorema de Bayes mostra que – como no exemplo da Guarda Costeira dos Estados Unidos –, à medida que a evidência se acumula, quaisquer que tenham sido os palpites iniciais, a

tendência é de eles se tornarem cada vez menos importantes, pois a evidência “fala por si mesma”.⁷

Os métodos frequentistas se tornaram cada vez mais populares, e alguns estatísticos tentaram advertir acerca dos perigos de varrer tudo isso para baixo do tapete. Eles foram mais ou menos ignorados durante décadas. Mesmo hoje, muitos pesquisadores continuam a usar métodos frequentistas para extrair conhecimento dos dados. Como consequência, inúmeras alegações, em campos que vão da economia e da psicologia à medicina e à física, na melhor das hipóteses, são questionáveis – e talvez estejam absolutamente erradas. A evidência disso agora começa a aparecer, e os pesquisadores lutam para replicar “descobertas” baseadas na lógica falha do frequentismo. Examinaremos essa questão perturbadora adiante, porém o aspecto mais chocante disso talvez seja o fato de que as falhas do frequentismo foram toleradas por muito tempo. Mas a coisa está mudando. O que hoje se conhece como métodos bayesianos está aos poucos sendo posto em uso pelos pesquisadores numa grande quantidade de campos. Isso em parte acontece pela sua potência. Até pouco tempo atrás, o arsenal completo de ferramentas bayesianas não era acessível para pesquisadores que – como o próprio Bayes – encontravam dificuldade de executar as somas necessárias para aplicar às questões da vida real. Isso agora foi resolvido pelo surgimento do poder de computação barato e abundante, permitindo que os métodos de Bayes sejam empregados em problemas muito sofisticados envolvendo uma porção de teorias concorrentes.

Ao mesmo tempo, os pesquisadores tomam cada vez mais consciência das muitas virtudes da fórmula milagrosa do reverendo Bayes. E, como veremos em breve, ninguém precisa fornecer números para tirar proveito deles.

Conclusão

As leis da probabilidade não se aplicam somente a triviais eventos do acaso, como lançar moedas. Podem também ser usadas para captar as noções geralmente indistintas de crença e evidência, e combiná-las de modo a produzir conhecimento novo. Chave para o processo é o teorema de Bayes, por muito tempo controverso, porém visto cada vez mais como a melhor maneira de dar sentido a uma evidência.

21. O encontro do dr. Turing com o reverendo Bayes

EM ABRIL DE 2012, o quartel-general da comunicação do governo do Reino Unido, o GCHQ (de Government's Communication Headquarters), finalmente revelou um dos seus segredos mais bem-guardados. O segredo assumia a forma de um documento técnico de 44 páginas descrevendo detalhes de um método espantosamente poderoso de decifrar códigos inimigos. É possível avaliar quão poderoso é o método pelo fato de o documento ter sido escrito durante a Segunda Guerra Mundial. Mas foram necessários mais de setenta anos antes que, como disse alguém do GCHQ, os matemáticos finalmente conseguissem “espremer seu sumo”. A liberação de documento tão secreto já era em si bastante notável, ainda mais quando se soube o nome de seu autor: o dr. Alan Turing (1912-1954), o brilhante matemático de Cambridge que desempenhou papel hoje celebrado na quebra dos códigos nazistas, seguindo depois adiante para se tornar pioneiro na criação dos computadores.

A mídia naturalmente deu grande importância à autoria de Turing de *The Application of Probability to Cryptography*. Para os conhecedores, porém, havia algo ainda mais impressionante no documento, com suas menções a evidências, probabilidades e o emprego de evidência a priori. Era a confirmação final daquilo de que se desconfiava pelo menos desde a década de 1970: que Turing e seus colegas do centro aliado de decifração de códigos em Bletchley Park haviam feito uso extensivo do teorema de Bayes. Os primeiros indícios de seu papel central vieram à tona num artigo sobre o trabalho estatístico de Turing na época da guerra, publicado em 1979 pelo seu colega matemático em Bletchley e entusiasta de Bayes, I.J. “Jack” Good.¹ Até a releitura de ideias bayesianas foi suficiente para provocar furiosas denúncias de emprego da crença a priori por parte de estatísticos preeminentes. Contudo, hoje está claro que, enquanto os estatísticos famosos empreendiam uma guerra intelectual contra Bayes e todos

os seus trabalhos no mundo externo, Turing e seus colegas do GCHQ os usavam em grande segredo para levar a vitória ao mundo mais que objetivo da Segunda Guerra Mundial e aos conflitos que se seguiram.

Quando estudantes, Turing e Good tinham conseguido contornar os dogmas frequentistas que então assolavam a comunidade de pesquisa. Eles se voltaram para o teorema de Bayes simplesmente porque ele parecia ideal para o cerne do desafio de quebrar códigos: transformar pistas e indícios em conhecimentos. Ao empregá-lo, eles trabalharam de trás para a frente, a partir dos dados observados – sinais inimigos interceptados –, para deduzir os esquemas mais prováveis de dispositivos de encriptação como a máquina Enigma, usada pelas forças armadas nazistas para criptografar suas comunicações operacionais. As engrenagens e conexões eram capazes de embaralhar mensagens em 15 bilhões de bilhões de maneiras diferentes, levando o chefe de Bletchley Park a manifestar suas dúvidas sobre se as mensagens seriam algum dia decifradas. Ele havia considerado essa possibilidade sem levar em conta o poder do teorema de Bayes para pegar até pistas frágeis, combiná-las com os dados e repetir o processo vezes e mais vezes, até surgirem os esquemas corretos – e as mensagens se tornarem legíveis.

Os decifradores de códigos de Bletchley em seguida turbinaram com Bayes o primitivo computador eletrônico, o Colossus, e usaram a combinação para decifrar a máquina Lorenza, muito mais exigente e que o próprio Hitler empregava para suas comunicações secretas com os comandantes de campo. Depois da guerra, é provável que tenham posto Bayes para trabalhar no maior triunfo do Ocidente durante a Guerra Fria, o Projeto Venona. Uma mancada da inteligência soviética na época da Guerra Fria introduziu uma minúscula falha no sistema de código usado pelos seus principais espões. Ainda não se sabe exatamente como isso foi explorado; contudo, a evidência sugere que, mais uma vez, Bayes e os computadores desempenharam aí algum papel. Na época em que o Projeto Venona foi encerrado, em 1980, ele havia desmascarado os mais famosos espões da Guerra Fria, entre eles Klaus Fuchs, Alger Hiss e Kim Philby.

Depois de voltar para a academia na década de 1960, Good tornou-se um dos raros a se dedicar aos métodos bayesianos durante a longa permanência dessa teoria na obscuridade. Às vezes era obrigado a assistir a palestras que faziam pouco dos métodos de Bayes, impedido, pelas regras de confidencialidade, de recorrer às suas próprias experiências para refutá-las.² O segredo intenso se justificava. Em 1951, dois matemáticos trabalhando no Venona para o serviço de informações dos Estados Unidos tiveram a permissão de publicar um artigo incorporando ideias bayesianas – e seu valor foi imediatamente notado pelos decifradores de códigos soviéticos.³

Eles adorariam pôr as mãos no documento ultrassecreto de Turing, uma verdadeira cartilha sobre a aplicação do teorema de Bayes ao problema geral da quebra de códigos. Partindo de princípios básicos, Turing aplica o teorema a sistemas cada vez mais complexos, dando exemplos trabalhados à medida que avança. A maior parte é bastante difícil de entender, mas há duas características no uso feito por Turing do teorema de Bayes encerrando lições que vão muito além do mundo secreto dos decifradores de códigos. A primeira é a indiferença que demonstrou ao se confrontar com o supostamente perigoso “problema dos a priori” – isto é, o estabelecimento de um nível de crença inicial a partir do qual começar o processo bayesiano de atualização utilizando nova evidência. Turing não tinha escrúpulos de resolver o problema com uma mistura judiciosa de fatos concretos e palpites bem-informados. Essas práticas eram consideradas anátemas pela maioria dos estatísticos influentes na época (e ainda hoje despertam controvérsias).

Felizmente para os Aliados, essa influência não se estendeu a Bletchley. Mesmo que o tivesse feito, é improvável que detivesse Turing, já renomado pelo pragmatismo e o desprezo pela autoridade. Como ele mostrou, contanto que os palpites estimativos iniciais não fossem absurdos demais, o teorema de Bayes tornava-os progressivamente irrelevantes à medida que chegavam novas evidências, resultando em informações úteis. Uma prova da alegação de Turing não poderia ser mais impressionante: a quebra do sistema de código dos inimigos que acelerou a derrota das potências do Eixo e salvou milhões de vidas.

Ironicamente, a reabilitação dos métodos bayesianos depois da guerra poderia ter sido mais rápida se eles não tivessem tanto êxito em aplicações muito vitais – e, consequentemente, secretas.

Mas há outro aspecto do emprego dado por Turing ao teorema de Bayes que o torna acessível mesmo para a maioria dos não matemáticos. Apesar de seu brilhantismo, nem Turing nem qualquer um de seus colegas se deleitava com cálculos complexos desnecessários. Como nós, eles achavam a soma mais fácil que a multiplicação, e isso os levou a reelaborar o teorema de Bayes numa fórmula mais fácil de usar, até mais intuitiva que o formato original, conservando, porém, toda a sua potência.⁴ (Ver Box a seguir.)

A VERSÃO DO DR. TURING DO TEOREMA DE BAYES

Novo nível de crença na teoria =

Velho nível de crença + Peso da evidência,

onde o *Peso da evidência* depende da chamada *Razão de probabilidade* (RP), a razão entre duas probabilidades condicionais: as chances de obter os dados observados presumindo-se que a crença esteja correta divididas pelas chances de obtê-los se a crença estiver errada. Isto é:

$$RP = \frac{\text{Pr (dados/crença correta)}}{\text{Pr (dados/crença errada)}}$$

A primeira coisa a notar em relação a essa pequena receita é que agora ela espelha exatamente a maneira como falamos sobre evidência e crenças. Doravante, temos uma fórmula na qual os dados são transformados em peso de evidência, que se soma ao nosso nível de crença. Ao escrever a fórmula dessa maneira, pegamos o modo mais básico de captar nossas crenças – como probabilidades variando de 0 a 1, passando por 0,5 – e as transformamos nas chamadas chances logarítmicas, que se estendem de menos infinito, numa extremidade, passam pelo zero e chegam a mais infinito, na outra ponta. Logo, a

forma como a força da crença é abrangida estende-se do ceticismo implacável, num extremo, até a certeza absoluta, no outro, passando por nem um nem outro, no meio – uma medida natural e lindamente simétrica da nossa crença. Ao mesmo tempo, ao contrário dos valores de zero e um da probabilidade, os equivalentes logarítmicos de menos e mais infinito servem para nos advertir de quão extremados são esses níveis de crença. Ou seja, eles ressaltam o fato de que ceticismo implacável e certeza absoluta não têm lugar no mundo real. Conectados ao teorema de Bayes, mostram também a irracionalidade de sustentar níveis de crença tão extremados que não podem ser alterados por nenhuma quantidade de evidência. Logo, essa formulação do teorema de Bayes não só capta a noção aparentemente inefável de crença, mas também mostra como mudá-la à luz da evidência – ao mesmo tempo que nos alerta para a inutilidade de aspirar a uma certeza do tipo divina. O próprio reverendo Bayes a teria aprovado.

Quanto à maneira como a nova evidência deve afetar nossas crenças, isso também se alinha melhor ao nosso senso comum. Como mostra o Box anterior, podemos somar, ou subtrair, peso de evidência ao nosso nível de crença. Quanto somar ou quanto subtrair, isso é ditado por um cálculo simples: pegamos as chances de observar os dados se nossa teoria for *verdadeira* e as dividimos pelas chances de observá-los se nossa teoria for *falsa*. Como seria de esperar, se a evidência for mais provável com a premissa de a teoria ser verdadeira do que ela ser falsa, isso *soma* peso de evidência em favor da nossa teoria. Por outro lado, se a evidência for mais provável com a premissa de a nossa teoria ser falsa, então ela *subtrai* peso de evidência. Mas há também uma terceira possibilidade que não devemos ignorar: que a evidência seja igualmente provável, independentemente de a teoria ser verdadeira ou falsa. Como o Box anterior mostra, isso leva a uma “razão de probabilidade” igual a 1, que o artifício do logaritmo transforma num peso de evidência igual a 0. Em outras palavras, essa evidência faz diferença zero no peso de evidência a favor ou contra a teoria ser verdadeira.

Tudo então é resumido no conjunto de regras intuitivas sobre avaliar o peso de evidência a favor ou contra uma crença específica (ver Box a seguir).

Temos agora o que precisamos para atualizar o nosso nível de crença numa teoria. E a “teoria” em questão não precisa ser alguma explicação esotérica de, digamos, forças subatômicas ou a origem do Universo (embora Bayes possa ser usado para tais coisas); pode ser *qualquer* hipótese, desde se um esquema particular foi usado para codificar uma mensagem específica da máquina Enigma até se novas evidências apoiam a crença em telepatia. O teorema de Bayes não se importa: qualquer que seja a noção que queiramos avaliar, ele nos diz o que precisamos considerar para criar peso de evidência a favor ou contra, e então que impacto isso tem sobre o nosso nível de crença. Na verdade, ele redundava numa fórmula de uma linha para “Como dar sentido à evidência”.

COMO DAR SENTIDO À EVIDÊNCIA

Para avaliar o impacto que a evidência deve ter sobre nosso nível de crença em alguma teoria ou alegação, precisamos saber (ou pelo menos ter palpites sobre) duas probabilidades: as chances de obter a evidência presumindo que nossa teoria esteja certa (vamos chamá-las de C) e as chances de obter a evidência presumindo que nossa teoria esteja errada (E). Então, o teorema de Bayes mostra que:

1. Se C for **maior** que E, temos um peso de evidência **positivo** que se soma à nossa crença de que a teoria está correta.
2. Se C for **menor** que E – isto é, a evidência é *menos* provável de surgir se a teoria estiver certa do que se a teoria estiver errada –, o peso de evidência é **negativo** e deve enfraquecer as nossas crenças.
3. Se C for **igual** a E, a evidência é igualmente provável *independentemente* de a teoria estar certa ou não; ela fornece peso de evidência **zero** às nossas crenças, e não devemos levá-la em consideração.
4. Se não temos (e não conseguimos nem adivinhar) C ou E, não podemos saber se a evidência é mais ou menos provável se a nossa teoria estiver certa – e devemos ter cautela para chegar a algum julgamento, *qualquer que seja*.

As duas primeiras regras do Box anterior são mais úteis quando temos alguns números concretos para empregar. Por exemplo, durante seu trabalho de quebra de códigos, Turing e seus colegas lançaram mão da sofisticada teoria da probabilidade para estimar as chances de obter relances de texto legível de que dispunham, com a premissa de terem os esquemas corretos da máquina codificadora inimiga, ou por puro lance de sorte. Aí somaram esse novo peso de evidência ao nível de crença existente presumindo terem os esquemas certos – resultando numa “quebra” completa.

A terceira e quarta regras, em contraste, são frequentemente úteis para testar alegações mesmo na ausência de números concretos. Tomemos, por exemplo, o argumento de que é possível descobrir se as pessoas gostam de manteiga segurando uma flor botão-de-ouro sob o queixo e atentando para o surgimento de um significativo brilho amarelo.^j Essa é uma ideia adorável e há muito usada pelos pais para distrair os filhos nos dias de verão – e as crianças a experimentam com amigos e descobrem a espantosa confiabilidade do teste. Todavia, a maior parte dos adultos sabe que há alguma coisa não muito correta na aparente confirmação do teste. O teorema de Bayes cristaliza essas dúvidas, porém vai mais longe, esclarecendo a regra básica para testar *qualquer* teoria. Como veremos, é uma regra que de hábito surpreende as pessoas espertas.

As suspeitas em relação ao “teste do botão-de-ouro” estão centradas no fato de que, como a maioria das pessoas gosta de manteiga, as chances de o teste dar positivo são muito altas, mesmo que ele não passe de um absurdo. O teorema de Bayes confirma essas suspeitas. A Regra 3 nos adverte que, se as chances de obter evidência são igualmente prováveis *independentemente* de a alegação ser verdadeira ou não, então o peso de evidência fornecido é zero. Logo, enquanto nossos filhos ficam impressionados vendo um brilho amarelo sob o queixo de todo mundo que gosta de manteiga (pelo menos nos dias de sol), o teorema de Bayes mostra que essa é só metade da história. O teste só pode gerar peso real de evidência se os resultados positivos não forem apenas *mais prováveis* com aqueles que gostam de manteiga, mas também *menos prováveis* com aqueles que não gostam – e isso impõe que façamos o teste com os *dois* tipos de pessoas. A

exigência de testes de comparação é muitas vezes desprezada pelos adultos, para não dizer pelas crianças, e é ressaltada pela Regra 4, ainda mais fácil de usar e mais amplamente aplicável. Por exemplo, se ouvimos relatos sobre algum impressionante teste novo para uma condição médica, precisamos saber mais do que apenas se o teste deu resultados positivos para pacientes portadores dessa condição (os chamados “verdadeiros positivos”). Para que o teste produza peso de evidência útil, também é necessário que se façam testes de comparação, a fim de verificar se ocorrem resultados positivos com pacientes não portadores da condição (os chamados “falsos positivos”). Sem isso, diz a Regra 4, devemos ter absoluta cautela em relação a qualquer julgamento acerca do valor do teste.

Mesmo que os pesquisadores tenham feito tudo isso, o teste diagnóstico é útil ao somar peso de evidência somente a um nível de crença já existente – e o teorema de Bayes mostra que, se esse nível era muito baixo (porque, digamos, a condição é muito rara), depois de somar o robusto peso de evidência, o nível de crença atualizado permanece muito baixo. Claro que o teorema será mais poderoso se pudermos inserir números para obter uma resposta quantitativa (como fizemos no Capítulo 20), mas o sentido já foi transmitido: não devemos ficar exageradamente impressionados com argumentos que se baseiam só em impressionantes taxas de verdadeiros positivos: precisamos mais que isso.

Quando temos os resultados, eles podem mudar o curso da história – como demonstraram Turing e seus colegas. Felizmente, não foi necessário esperar pela liberação de seu relatório para que a potência do teorema de Bayes fosse reconhecida de maneira mais ampla. Sua capacidade de quantificar o processo central da ciência – atualizar conhecimento à luz de nova evidência – tem encontrado utilidade numa gama de campos cada vez maior. Os médicos que testam uma nova terapia exploram a capacidade de combinar o conhecimento existente com dados novos, o que lhes permite chegar mais depressa a uma conclusão sobre a eficácia, com maior confiabilidade e usando menos pacientes.⁵ Os paleontologistas que tentam desvendar a evolução do *Homo sapiens* empregam métodos bayesianos para comparar teorias rivais e se concentrar nas

mais plausíveis,⁶ enquanto os cosmólogos os utilizam a fim de determinar as propriedades do Universo com uma precisão sem precedente.⁷

O teorema de Bayes também tem uma miríade de usos menos reconhecidos, mas não menos impressionantes, acelerando nossas pesquisas on-line, corrigindo nossos erros de digitação e protegendo-nos de todos aqueles e-mails indesejados, pela capacidade que tem de aprender a partir do que já é sabido. Num maravilhoso exemplo da repetição da história, o teorema utilizado por Turing e seus colegas com efeito tão triunfal durante a Segunda Guerra Mundial é empregado agora contra um novo inimigo global: os cibercriminosos.

De impérios de mídia multinacionais a companhias petrolíferas, de firmas de defesa a sites de encontros, as redes de computadores estão agora sob constante ataque dos hackers. No ciberespaço equivalente à evolução darwiniana, toda contramedida é recebida com uma resposta cada vez mais sofisticada – e o crescente reconhecimento de que as velhas técnicas de senha e encriptação já não bastam. Muitos ataques hoje são perpetrados por gente de dentro capaz de contornar os sistemas de segurança. Todavia, há uma coisa que nunca muda em relação aos cibercriminosos: por definição, eles estão atrás de informação sensível. Não importa quanto finjam querer outra coisa, acabarão por revelar suas verdadeiras intenções – fuçando arquivos pessoais, por exemplo, ou tentando baixar dados. Em suma, como seus correlatos no mundo real, os cibercriminosos têm um *modus operandi*, “MO”, que pode ser apreendido e procurado.

Identificar essa atividade atualmente é encarado como algo vital na luta contra o cibercrime. Liderando esse ataque está uma empresa sediada na Grã-Bretanha, conhecida como Darktrace. Grande parte de seu pessoal é formado no GCHQ, o equivalente moderno de Bletchley Park, onde Turing e seus colegas realizaram seus milagres. A força motriz por trás da estratégia da Darktrace é um método para descobrir como é o aspecto de uma rede de computadores quando ela funciona bem, revelando assim quando funciona mal. E no seu coração está nada mais que a milagrosa fórmula do reverendo Bayes.

Conclusão

Mesmo na ausência de números concretos, o teorema de Bayes ajuda a revelar exatamente que perguntas devemos fazer acerca da evidência. E também nos alerta para quando estão nos contando apenas a metade do que precisamos saber – e às vezes ainda menos.

^j Esse “teste” é feito em países de língua inglesa; ele é possível pelo nome do botão-de-ouro (ou ranúnculo) em inglês: *buttercup*, que significa literalmente “caneca (ou copo, xícara) de manteiga”; se o brilho amarelo aparece quando se segura a flor sob o queixo, então a pessoa gosta de manteiga. (N.T.)

22. Usando Bayes para julgar melhor

POR VOLTA DAS QUATRO HORAS da manhã do dia 21 de julho de 1996, depois de horas de interrogatório pelos detetives de Luisiana, Damon Thibodeaux finalmente sucumbiu e confessou o assassinato da prima. O corpo dela havia sido encontrado no dia anterior nas margens do Mississippi, e Thibodeaux não hesitou em revelar tudo o que fizera: como a golpeara no rosto, a estuprara e finalmente a estrangulara com um arame que estava em seu carro. O julgamento durou apenas três dias, e o júri levou menos de uma hora para dar o veredito: culpado. Ele foi condenado à morte por assassinato agravado por estupro.

Thibodeaux passou os quinze anos seguintes no Corredor da Morte, até que finalmente, em setembro de 2012, foi inocentado de todas as acusações e libertado. Ele se tornara a 300^a pessoa nos Estados Unidos a se provar inocente a partir da evidência do DNA – mas era somente o último dos incontáveis milhares que, ao longo dos séculos, haviam sido condenados com base em evidências inconsistentes.

Após sua libertação, Thibodeaux explicou como chegara a acreditar que tinha cometido o crime: uma mistura de privação de sono, pressão implacável e um avassalador desejo de que tudo simplesmente acabasse. Mesmo durante o julgamento estava claro que sua “confissão” se baseava numa mistura de indícios tirados das alegações dos detetives e pura invenção. A vítima fora golpeada com um instrumento não cortante, e não com a mão, estrangulada com um arame tirado de uma árvore, e não do carro – e não havia evidência de atividade sexual, forçada ou não. Thibodeaux chegou a dizer aos interrogadores: “Eu não sabia que tinha feito, mas fiz.”

Em suma, esse foi um caso clássico de falsa confissão, servindo apenas para adicionar peso à crença de que essa antiquíssima forma de “evidência” é mais frágil que o papel na qual está escrita. Ninguém sabe disso melhor que os

membros do Innocence Project, criado em 1992 na Escola de Direito Cardozo, de Nova York, para reexaminar aparentes erros da Justiça. Na época em que escrevo este livro, o trabalho do grupo já inocentou trezentas pessoas de crimes sérios pelos quais foram condenadas e que não cometeram, tendo sido obrigadas a cumprir bem mais de uma década de cadeia – e, como Thibodeaux, muitas delas no Corredor da Morte. Mais de um quarto das condenações injustas revertidas pelo Innocence Project envolvia confissões falsas. E mal se resiste a pensar nessa taxa de erro em países com menos consideração pelos devidos processos legais.

Muitos de nós temos uma desconfiança inata na evidência confessional – e essa é uma atitude respaldada pelas implicações mais básicas do teorema de Bayes. Como vimos no capítulo anterior, para qualquer fonte de evidência adicionar peso às nossas crenças sobre uma teoria, uma condição muito específica deve ser aplicada. E para que uma confissão adicione peso de evidência à nossa crença na culpa da pessoa, essa condição é:

$\text{Pr}(\text{confissão, dada culpa})$ deve *exceder* $\text{Pr}(\text{confissão, dada inocência})$

Falando sem rodeios, devemos ter confiança de que as chances de obter uma confissão do culpado sejam maiores que as chances de obter uma confissão do inocente. Claro, pode-se debater isso – e o ponto é precisamente este: é óbvio que a confissão não pode ser considerada inquestionável em todos os casos. De fato, colocando-se na situação de Thibodeaux, a pessoa se vê pressionada além de toda e qualquer resistência até que esteja disposta a falar qualquer coisa – a única dúvida é quanto tempo isso levará. Para alguns, a confissão pode exigir tortura extrema; para outros, a mera possibilidade de quinze minutos de fama na TV já se mostrou motivação suficiente. O teorema de Bayes deixa clara a restrição imposta às duas possibilidades – e o fato de que ela está longe de uma garantia absoluta. Na verdade, uma pensada rápida sugere que, para a maioria dos crimes, o que se sustenta é exatamente o contrário. Por exemplo, se foi cometido um crime no mundo das gangues, é possível ter uma boa certeza de que o perpetrador é o assassino profissional de alguma gangue. Será que essas

pessoas, com seus códigos de *omertà*, realmente têm maior probabilidade de confessar durante os interrogatórios do que um inocente levado para ser interrogado? Lembre-se de que Bayes mostra que não basta que tais pessoas tenham *a mesma probabilidade* de confessar; elas precisam ter *maior probabilidade* de confessar para que sejam fontes úteis de evidência de culpa em tais casos.

Essas dúvidas ficam ainda mais fortes nos crimes relacionados a terrorismo, quando se sabe que os perpetradores em geral são treinados para resistir aos interrogatórios. Agora temos uma situação em que os culpados de ataques terroristas de fato têm *menos* probabilidade de se dobrar ao interrogatório que uma pessoa inocente. Nesse caso, Bayes nos diz algo bem chocante: o simples fato de que alguém acusado de um ato terrorista tenha confessado torna o suspeito *menos* provável como perpetrador. Bayes nos diz que os condenados por esses crimes com base em confissão podem muito bem ser vítimas de erros judiciais. Talvez não seja coincidência o fato de que a prova confessional de supostos terroristas ocupe grande espaço em alguns dos mais importantes erros da Justiça em muitos países, como os casos dos Quatro de Guildford e dos Seis de Birmingham, no Reino Unido na década de 1970.¹

Claro que em muitos casos há uma evidência mais convincente que apenas a confissão, fornecida por fontes de peso de evidência adicionais e mais confiáveis – como as baseadas na ciência forense. Pelo menos, seria bacana pensar assim. O problema é que uma quantidade grande demais de testes científicos forenses tem sido aceita na corte sem passar pelo “crivo bayesiano” para conferir se realmente eles adicionam peso de evidência.

Peguemos o conhecido caso dos Seis de Birmingham, em 1975, no qual seis homens foram condenados pelo ataque do IRA a dois pubs em Birmingham, em atentados que mataram 21 pessoas e feriram mais de 180. Quatro dos seis homens assinaram confissões logo depois de presos, mas não foi apenas a evidência confessional que selou sua sorte. Três dos quatro também deram positivo no chamado teste de Greiss para contato com explosivos. Segundo o

cientista forense, o resultado foi tão forte que ele tinha “99% de certeza” de que alguns dos réus haviam manuseado explosivos.

Não fica muito óbvio o que ele queria dizer com isso; o mais provável é que estivesse se referindo ao fato de que o teste é bastante efetivo para detectar traços de nitritos da nitroglicerina. O problema é que, como mostra o teorema de Bayes, saber que uma fonte de evidência tem alta probabilidade de dar positivo nas circunstâncias certas (ou seja, ter alta taxa de “verdadeiro positivo”) é só metade da história; para estabelecer seu peso de evidência, também precisamos da taxa de falsos positivos – e, mais ainda, esta precisa ser *mais baixa* que a taxa de verdadeiros positivos. Esse fato essencial nunca foi sugerido no julgamento. Surpreendentemente, só em 1986 – mais de uma década após a condenação – os cientistas forenses do governo do Reino Unido analisaram o relatório sobre a questão. Descobriram que o teste de Greiss era bem capaz de dar resultados positivos quando aplicado a mãos de pessoas que tinham jogado cartas ou que não as tivesse lavado depois de urinar. Em outras palavras, o teste tinha uma taxa impressionante de verdadeiros positivos, mas também uma taxa significativa de falsos positivos, solapando o peso da evidência.² Tampouco era a primeira vez que se arrolavam dúvidas em relação a esses testes: a mesma questão fora identificada uma década antes de os Seis de Birmingham irem a julgamento, com falsos positivos gerados por teste similar usado nos Estados Unidos desde os anos 1930.

Muito mais preocupante, porém, é o fato de que ainda há testes forenses que não passaram por um crivo bayesiano adequado, resultando na prisão de pessoas inocentes. Segundo o Innocence Project, quase metade dos trezentos e tantos casos de erros judiciais revelados envolve testes forenses mal interpretados, mal aplicados e nunca adequadamente validados. Mesmo técnicas conhecidas e amplamente usadas, como microscopia capilar, análise de solas de calçados e comparações de mordidas dentárias, jamais passaram pela peneira bayesiana para avaliar que peso de evidência – havendo algum – podem fornecer.

Em contraste, o Innocence Project tem uma profusão de evidências de fracassos, como no caso de Steven Barnes, condenado pelo estupro e assassinato

de uma mulher em Whitestown, Nova York, em 1989. Para o júri, essa ocorrência parecia inquestionável. Embora os relatos de testemunhas oculares fossem duvidosos, a evidência forense era avassaladora. O tipo de solo encontrado nos pneus do caminhão de Barnes tinha características semelhantes às da cena do crime, e o padrão do tecido dos jeans da vítima tinha os mesmos traços que a marca que ficara no caminhão. E, talvez o mais significativo de tudo, o exame microscópico de dois fios de cabelo encontrados no caminhão tinha traços diferentes do de Barnes, contudo, mais uma vez, eram similares aos da suposta vítima. Outros testes se mostraram inconclusivos, mas o júri já tinha ouvido o suficiente, e Barnes foi condenado à pena mínima de 25 anos. Ele foi um dos primeiros casos do Innocence Project, e a equipe identificou uma legião de falhas na acusação. Entre essas falhas estava o fato de que os testes de cabelo, de semelhança de tecido e de solo nunca haviam sido cientificamente validados.

Barnes foi finalmente inocentado em 2009, quase vinte anos depois da condenação. Nesse mesmo ano, a ciência forense viu-se na berlinda da respeitabilidade científica, com ninguém menos que a academia Nacional de Ciências dos Estados Unidos no comando da acusação. Num relatório intitulado *Strengthening Forensic Science in the United States*, a academia não mediu palavras: os dados exigidos pelo teorema de Bayes para estabelecer o peso de evidência “são componentes-chave da missão da ciência forense”, e declarações explícitas e precisas são “absolutamente importantes”.

Felizmente, para os casos semelhantes ao de Barnes, Thibodeaux e de muitas outras pessoas inocentes, existe um teste forense com base científica sólida e taxas bem-estabelecidas de verdadeiros e falsos positivos: o perfil de DNA. Desde que foi usado pela primeira vez em 1987 (por acaso, para inocentar alguém que havia confessado falsamente um duplo assassinato na Inglaterra), o teste não só tem ajudado a capturar inúmeros criminosos, como também revelou as falhas de muitos testes forenses supostamente “científicos”. O perfil de DNA tornou-se o padrão-ouro ao qual recorreram o Innocence Project e muitos outros em busca da verdade. Contudo, o teorema de Bayes mostra que até o teste de

DNA pode se ver minado pela falha em compreender o processo pelo qual a evidência se torna conhecimento.

O teste de DNA deve sua reputação ao fato de que todo mundo, exceto gêmeos idênticos, tem um perfil genético único, empacotado na famosa molécula da dupla-hélice espremida nas células. Isso confere à técnica uma taxa altíssima de verdadeiros positivos: é virtualmente certo que o DNA encontrado numa cena de crime combinará com aqueles que lá estavam – incluindo o culpado. Portanto, ele tem uma taxa de verdadeiros positivos de quase 100%. Mas, como sempre, Bayes nos adverte para não ficarmos impressionados demais com esse fato; devemos saber as chances de se obter uma combinação com alguém que não esteve na cena do crime – a taxa de falsos positivos. O teorema de Bayes mostra que, quanto maior a diferença entre as taxas de verdadeiros e falsos positivos, maior o peso da evidência. A cifra exata depende da qualidade do DNA e de quantas “combinações” são encontradas com a amostra tirada do suspeito. Resumindo, pela natureza química do DNA, não é incomum obter muitas combinações numa amostra, reduzindo a taxa de falsos positivos para valores baixíssimos, de até 1 em vários milhões.

Agora temos os dois componentes necessários para o peso de evidência, e os perfis de DNA claramente fornecem grandes quantidades de evidência. Inserir os números no teorema de Bayes mostra que a técnica pode elevar as chances a priori num fator de vários milhões. Mas também deixa claro que ainda precisamos saber quais são essas chances a priori antes de chegarmos a alguma conclusão sobre a culpa ou inocência do acusado. Se houver muito pouco de outras evidências, esse nível a priori será extremamente baixo. Por exemplo, se só sabemos, antes do teste de DNA, que o criminoso era um homem inglês, o nível de crença a priori de que o suspeito seja culpado é de apenas 1 em 30 milhões – a população masculina da Inglaterra. Assim, mesmo depois de ser ampliado por um fator de vários milhões, ainda podemos acabar com um nível de crença de mais ou menos 1 em 10 – que ainda está muito longe de ser “além de qualquer dúvida razoável”. Apesar de tudo isso, e do perigo claro e sempre presente de uma interpretação errada, a evidência de DNA ainda é rotineiramente

apresentada sem nenhuma referência a Bayes, deixando claro o que significa essa evidência e como incorporá-la às outras para chegar a um veredito final. Os jurados tentam dar sentido a declarações de cientistas forenses como: “A probabilidade de obter uma combinação de DNA tão boa de alguém que não esteja ligado à cena do crime é de 1 em 3 milhões.” Sem Bayes para ajudar a deixar claro que este é apenas um enunciado da taxa de falsos positivos, há um alto risco de confundi-la com as chances de o suspeito ser inocente, o que, com 1 em 3 milhões, aparentemente implica culpa além da dúvida razoável.

Considerando o papel central das evidências nos tribunais, e o fato de o teorema de Bayes dar sentido a elas, há uma clara necessidade de que a pessoa lidando com a evidência forense tenha alguma ideia de suas implicações. Simplesmente estar cômico das regras que governam o peso da evidência já basta para evitar as armadilhas mais óbvias quando se avalia uma evidência. No entanto (o que é incrível), no Reino Unido o Judiciário rejeitou especificamente essa modesta proposta. Numa decisão de 1997 – amplamente condenada na época e ainda objeto de muito debate –, a Corte de Apelação inglesa determinou “não ser apropriado” que os jurados deem sentido à evidência usando “fórmulas matemáticas como o teorema Bayes [sic]”, pois isso usurparia a tarefa do júri, de pesar toda evidência em conjunto. De fato isso aconteceria, mas provavelmente também iria fazer com que se confiasse menos nas evidências falhas, com que houvesse menos confusão quanto ao seu significado e menos erros judiciais.

Conclusão

É confortador pensar que os júris já não baseiam seus vereditos em julgamentos decididos por ordálios, boatos e falsas confissões. No entanto, o peso da evidência fornecida por muitos testes forenses supostamente “científicos” jamais foi estabelecido da forma adequada. Até que isso aconteça, esses testes continuarão a desempenhar importante papel nos gritantes erros judiciais.

23. Um escândalo de significância

COMO ACONTECE COM a maioria das publicações científicas acadêmicas, *Basic and Applied Social Psychology* não é uma revista famosa. Fundada em 1980, tem leitores especializados, circulação modesta e nada semelhante à influência de publicações de pesquisas de primeira linha como a *Science* ou a *Nature*. Mesmo assim, em 2015, a *Basp* conseguiu provocar controvérsia em círculos científicos quando seus editores declararam que não aceitariam mais conclusões de pesquisas baseadas em “testes de significância”.

Esse parece um daqueles assuntos que só os especialistas entendem ou com que se preocupam. Todavia, deveria preocupar a todos nós, porque os editores da *Basp* enfatizavam um aspecto que ameaça a confiabilidade da pesquisa científica. Ele está centrado nos métodos amplamente usados por pesquisadores para decidir se descobriram algo digno de ser levado a sério. Como uma espécie de teste quantitativo decisivo, esses métodos são aplicados a descobertas experimentais para saber se podem ser consideradas “estatisticamente significativas”. A questão tem importância fundamental, pois essas descobertas têm maior oportunidade de serem publicadas nas revistas de pesquisa, propiciando louros e verbas para os pesquisadores. Em alguns casos, podem gerar novas áreas de investigação, influenciar políticas públicas e até modificar práticas globais.

A questão é que há alguns problemas muito sérios nesse teste decisivo. Primeiro, o critério usado para chegar à significância estatística não é confiável, tendendo assustadoramente a produzir resultados fortuitos como se fossem efeitos genuínos. Segundo, ele é enganoso, incentivando aqueles que o empregam a acreditar que aquilo que descobriram é realmente “significativo”, no sentido de ser importante. No entanto, o mais inquietante de tudo é que muitos pesquisadores – talvez a maioria – não entendem realmente como e por

que seus resultados passaram no teste da significância estatística. Por conseguinte, uma fração substancial das incontáveis conclusões de pesquisas feitas ao longo de décadas com base na “significância estatística” não passa de um absurdo sem sentido.

A própria ideia de que gerações de cientistas vêm usando uma técnica furada para dar sentido às evidências é revoltante. Se isso fosse verdade, o fato já não teria sido apontado décadas atrás? E se o furo realmente fosse tão sério, haveria evidências de sobra de que muitas das descobertas de pesquisa estão minando o progresso científico, não? Na verdade, foi apontado, sim, e há, sim, evidências de sobra. Desde que foram inicialmente adotados há mais de oitenta anos para dar sentido à evidência científica, os testes de significância estão na mira de alguns dos mais eminentes estatísticos da época.¹ Mesmo seu inventor, o professor Ronald Fisher, da Universidade de Cambridge – amplamente considerado um dos fundadores dos modernos métodos estatísticos –, manifestou preocupações relativas a erros de interpretação. A cada tanto, revistas acadêmicas e sociedades de especialistas têm abordado a questão, ponderado por algum tempo, apenas para deixá-la de lado novamente. A recusa por parte da *Basp* dos testes de significância recebeu por um breve tempo manchetes na mídia acadêmica, mas também parece improvável que ela venha a provocar alguma mudança mais significativa.

Essa visível complacência é ainda mais difícil de entender dada a abundante evidência de que os testes de significância não são adequados para seu propósito. Durante anos, a maior parte das evidências eram anedóticas; a maioria delas na forma de estudos de questões de saúde nunca parecia chegar a um consenso, como seria razoável esperar, se houvesse algum efeito genuíno em andamento. Telefones celulares e câncer no cérebro, linhas de alta tensão e leucemia infantil, elos genéticos para todos os tipos de traços – a evidência fluía e escoava, sem qualquer sinal de resolução. Os estudos às vezes se contradiziam mutuamente com uma naturalidade quase ridícula; numa semana, uma revista famosa publicava um achado de pesquisa digno de manchetes, só para ser aparentemente desbancado logo em seguida.²

As explicações para esses fracassos são numerosas demais para se chegar a um consenso. Como vimos nos Capítulos 10 e 11, os estudos podem ser minados por uma série de fatores, como a falta de randomização. Podem ser pequenos demais para detectar algum efeito real, ou tão grandes que os pesquisadores têm alta probabilidade de chegar a algum resultado impressionante por mero acaso – basta que insistam em procurar.³ Tudo isso forneceu camuflagem conveniente para uma inconveniente verdade: os “testes de significância” podem fazer os dados mais triviais parecerem ouro científico.

A evidência disso tem se mostrado presente há décadas – para quem estiver disposto a enxergar. Em 1995, a renomada revista de pesquisa *Science* trouxe uma matéria especial⁴ relatando o que se chamaria “O curioso caso do importante avanço que sumiu”. O foco da reportagem era a epidemiologia, campo no qual os pesquisadores habitualmente ganham as manchetes argumentando que uma ou outra atividade, desde tomar café até usar panelas de alumínio, está relacionada a um ou outro efeito sobre a saúde, desde ataques cardíacos até doença de Alzheimer. Os estatísticos entrevistados para a reportagem da *Science* advertiam que todo o campo era vulnerável à difundida concepção errônea do real significado da expressão “significância estatística”. Ainda assim, suas preocupações pareciam picuinhas acadêmicas em comparação com a miríade de outras causas mais familiares de conclusões não confiáveis, como amostragens de tamanho inadequado e grupos de estudo mal escolhidos. Mesmo assim, o estranho sumiço de uma evidência supostamente gritante continuou na agenda da pesquisa, da psicologia à nutrição e à economia.

Uma década mais tarde, o distinto estatístico médico John Ioannidis, da Universidade de Stanford, publicou um celebrado artigo com o título: “Por que a maioria dos achados de pesquisa publicados são falsos”,⁵ frisando aquilo que muitos estatísticos vinham dizendo por décadas: a testagem de significância estatística é uma “estratégia conveniente, porém mal fundamentada” para se chegar a conclusões científicas. Suas implicações, de que mais de 50% de *todas* as descobertas científicas estão erradas, podem ser criticadas, na melhor das hipóteses, como não substanciadas e talvez um exagero crasso. Dito isso,

tentativas de avaliar a escala do problema replicando estudos publicados sugerem que cerca de 1 em cada 5 conclusões de pesquisa são falsos positivos, com um número ainda mais alto em algumas disciplinas.⁶ Dada a colossal quantidade de tempo, esforço e dinheiro (atualmente, por volta de US\$ 1,5 trilhão por ano, globalmente)⁷ gastos em pesquisa científica, se essas cifras estiverem remotamente corretas, elas representam um escândalo de proporções estarrecedoras.

O que há exatamente de errado com essas técnicas, concebidas e promovidas por um dos fundadores da estatística moderna, ainda ensinadas e servindo de base para pesquisadores no mundo todo? Por que estes relutam tanto em abandoná-las – e o que deveriam fazer em lugar delas? A esta altura, talvez não seja nenhuma surpresa saber que as respostas estão na receita de Bayes, com seus 250 anos, para dar sentido à evidência – e os problemas que os cientistas têm tido com suas implicações desde então.

A falha fundamental na maneira como a evidência científica é geralmente avaliada reside neste fato simples: como mostrou Bayes, não se podem pegar sem mais nem menos afirmações como “a probabilidade de A, dado B”, e invertê-las, gerando “a probabilidade de B, dado A”, e presumir que a resposta deve ser sempre a mesma. Claro que ela *pode* ser – se os eventos A e B forem independentes. Por exemplo, se estamos lançando uma moeda honesta, óbvio que não há problema em assumir isso porque:

$$\text{Pr}(\text{obter cara no segundo lançamento, dado que obtivemos coroa no primeiro}) = \frac{1}{2}$$

pode ser simplesmente invertido de modo a dizer que

$$\text{Pr}(\text{obter coroa no segundo lançamento, dado que obtivemos cara no primeiro}).$$

Isso também é igual a $\frac{1}{2}$, porque os dois eventos são independentes, então a ordem não importa. Mas em geral não podemos usar esse recurso, mesmo com eventos simples. Por exemplo, é evidente que, se estamos jogando cartas, seria uma loucura lançar mão desse argumento, porque sabemos que

$$\text{Pr}(\text{segunda carta tirada é menor que um ás, dado que a primeira carta foi ás})$$

é bastante alta (pois há muitas cartas), assim podemos inverter a afirmação e apostar pesado em tirar um ás na segunda carta, pois

$\text{Pr}(\text{segunda carta tirada é ás, dado que a primeira carta foi menor que ás})$

também deve ser bem alta. Os dois eventos, “primeira carta é X, segunda carta é Y”, claramente se afetam mutuamente, não são independentes – então, a ordem importa. Bayes nos deu os meios de trocar essas probabilidades “condicionais” de lugar em todas as circunstâncias, e – o que é mais importante – nos diz que, para fazê-lo, também precisamos saber as probabilidades incondicionais dos dois eventos. Até aqui, é simples; então, qual o grande problema? Ele vem quando começamos a usar probabilidades como medidas do nosso grau de crença em alguma coisa. O processo de troca de lugar pode levar a inferências que não são apenas tolas, mas perigosamente enganadoras. Preocupado com a dor de cabeça recorrente que está tendo, você entra na internet e descobre o perturbador fato de que os seus sintomas muitas vezes estão associados a tumores no cérebro, e que

$\text{Pr}(\text{ter dor de cabeça, dado que você tem um tumor cerebral})$

é cerca de 50-60%.⁸ Nesse ponto, é muito fácil inverter as coisas e concluir que

$\text{Pr}(\text{ter um tumor cerebral, dado que você tem dores de cabeça})$

também é cerca de 50-60%. Felizmente, porém, tendo lido este livro, você sabe que só pode fazer esse cálculo de modo confiável usando o teorema de Bayes, e isso exige que você leve em conta probabilidades a priori. Na verdade, botando para funcionar a versão completa do teorema, sabemos que

$\text{Chances}(\text{tumor cerebral, dadas dores de cabeça}) = \text{RP} \times \text{Chances}(\text{tumor cerebral}),$

onde RP é a razão de probabilidade, dada por

$\text{Pr}(\text{dores de cabeça, dado tumor cerebral})/\text{Pr}(\text{dores de cabeça, dada ausência de tumor cerebral}).$

Agora vemos que há muito menos motivo de preocupação, e por duas razões. Primeira, e mais importante: tumores cerebrais felizmente são incomuns, sendo

diagnosticados em cerca de uma entre vários milhares de pessoas por ano. A probabilidade a priori de que sejamos uma delas também é muito baixa – tornando Chances(tumor cerebral) igualmente baixas. Mas ainda assim poderíamos ter motivo para nos preocupar se essas chances a priori baixas fossem ampliadas por uma RP muito alta. Já temos metade da informação necessária para calcular isso: a cifra de 50-60% para as chances de ter dores de cabeça se tivermos um tumor no cérebro. Felizmente, porém, essa é só a parte de cima, o numerador da RP: também precisamos da probabilidade de ter dores de cabeça se *não* tivermos um tumor no cérebro. Como dores de cabeça são muito comuns, essa probabilidade também é muito alta, portanto, a RP não o é. Conclusão: chances a priori baixas combinadas com RP inexpressiva levam a chances baixas de tumor cerebral, dada a evidência de dores de cabeça. Logo, a lição é clara: sempre que quisermos saber

$\Pr(\text{nossa teoria estar certa, dada a evidência}),$

precisamos estar cientes de que podemos estar cometendo um erro enorme pensando em obter isso simplesmente trocando de lugar o valor de

$\Pr(\text{a evidência observada, dada a nossa teoria estar certa}).$

Ainda assim, incrivelmente, esse é o tipo de armadilha na qual os pesquisadores caem sempre que usam a significância estatística para decidir se estão fazendo uma descoberta interessante. Na verdade, é pior que isso – com consequências devastadoras para o progresso científico. Para verificar, é necessário fazer uma coisa que pouquíssimos pesquisadores fazem, e nos familiarizar com a probabilidade que está no núcleo do problema: o inocuamente denominado “valor p ”. Felizmente, isso não é difícil – embora as implicações não sejam nada felizes para o progresso da ciência.

Introduzido por Fisher em 1925, em seu famoso *Statistical Methods for Research Workers*, o valor p parece uma forma esmerada de avaliar o risco de um resultado científico ser apenas um sinal aleatório. Claro que nenhum cientista quer fazer muito alarde acerca de uma descoberta fortuita. Fisher sugeriu que um

modo de realizar essa avaliação era calcular o valor p , que definiu como as chances de obter resultados pelo menos tão surpreendentes quanto os obtidos, *presumindo* que eles realmente sejam só casualidades (ver Box a seguir).

**O MÉTODO DO VALOR P DO PROFESSOR FISHER:
ENGANANDO PESSOAS INTELIGENTES DESDE 1925**

1. Calcule o valor p para o resultado do seu estudo usando fórmulas para:

Pr(obter resultado pelo menos tão surpreendente quanto o observado, assumindo que ele é apenas fruto do acaso).

2. Se essa probabilidade for menor que 5%, chame o resultado de “estatisticamente significativo”.
3. Enuncie o resultado no seu artigo científico acompanhado do valor p , alegando que este dá sustentação à sua teoria.

Fisher veio com a seguinte regra, relacionando valores p com significância estatística: se o valor p calculado para um achado estiver abaixo de 5%, então o achado pode ser considerado “estatisticamente significativo”. Tudo isso soa muito bem, ainda que um pouco confuso. Mas há uma cilada enorme à espera daqueles que alegremente aceitam a sugestão. Fisher está dizendo que um resultado é estatisticamente significativo se as chances de obter um resultado pelo menos tão surpreendente, *presumindo* que ele seja realmente fortuito, estiverem abaixo de 5%. No entanto, por que haveria alguém de se incomodar com uma coisa dessas – e de onde vem esse valor de 5%? Não deveríamos enxergar algo bem menos enrolado, isto é, as chances de que os nossos resultados *realmente* se devam apenas ao acaso, calculando

Pr(resultados obtidos serem por acaso, dado o resultado obtido),

e então checar se *isso* está abaixo de 5%? Ou, então, que tal esquecer toda essa baboseira de resultados por acaso e simplesmente calcular

$\Pr(\text{resultados obtidos refletem algum efeito genuíno, dado o resultado obtido}),$

e ver se isso *excede* 95%? Será que *esta* não seria uma definição muito mais clara, intuitiva e relevante de resultado “significativo”? De fato, seria – e note como é diferente do que Fisher oferece. A alternativa focaliza os resultados reais obtidos, e não os estranhamente maquiados “resultados pelo menos tão surpreendentes”, e se esses resultados refletem um efeito genuíno, em vez de ser mais outra explicação rival – em outras palavras, que são mera casualidade. Contudo, o mais preocupante, a definição de Fisher de um valor p deve ser revirada para chegar perto do que realmente deveria interessar aos cientistas. Isto é, o valor p é calculado com a *premissa* de que a única explicação para os resultados é o acaso. Como tal, é claro que não podemos simplesmente revirar o valor p e alegar que esse mesmíssimo número agora representa as chances de que a premissa seja incorreta. Esse é o clássico furo da inversão, e é tão pouco confiável quanto presumir que, como há grandes chances de ter dor de cabeça, dado um tumor cerebral, há exatamente a mesma chance de um tumor cerebral, dadas as dores de cabeça.

No entanto, essa é a asneira que inúmeros pesquisadores vêm cometendo desde que Fisher apresentou pela primeira vez seu teste de “significância” com valor p , tantos anos atrás. E as consequências têm enchido as páginas das revistas de pesquisa desde então: resultados bizarros, que não seriam levados a sério nem por um momento se não tivessem passado pelo critério de “significância”.

Então, o que se apossou de Fisher para vir com uma definição tão estranha? Resumindo, sua determinação de evitar o que Bayes nos diz ser inevitável: a introdução de conhecimento e crenças a priori na interpretação de dados científicos. O professor Fisher era um matemático brilhante e sabia muito bem dos perigos de inverter probabilidades condicionais impunemente. E também conhecia tudo sobre o teorema de Bayes, o problema dos a priori que ele criava e como Bayes, Laplace e outros haviam tentado lidar com ele. Mas não quis saber daquilo tudo – muito menos de afastar as crenças subjetivas na avaliação das evidências. A repulsa de Fisher era visceral, embora tentasse com frequência

disfarçá-la usando razões técnicas aparentemente desapaixonadas para rejeitar os métodos de Bayes.⁹ Depois, ele não teve escolha a não ser inventar alguma medida não bayesiana para os pesquisadores usarem ao tentar dar sentido a suas descobertas. O resultado foi o valor p , cuja definição claramente maquinada reflete suas origens: a busca condenável de evitar o inevitável. Não é possível avaliar a probabilidade de um resultado, se este é fruto do acaso, apenas usando valores p . O emprego que Fisher faz do termo “significativo” para resultados com baixos valores p parece um artil semântico para se esquivar de um fato matemático. Com certeza criava-se o risco de os valores p serem mal interpretados, e foi o que aconteceu. De início, até o próprio Fisher caiu na armadilha de inverter valores p baixos e interpretá-los como baixa chance de o resultado ser fortuito. Justiça se faça, poucos anos após o aparecimento do seu livro-texto, Fisher advertiu sobre os perigos de excessos de interpretação do seu conceito de significância:

O teste de significância só diz a ele [ao investigador prático] o que ignorar, ou seja, todos os experimentos nos quais não se obtêm resultados significativos. ... Consequentemente, resultados significativos isolados que ele não sabe como reproduzir são deixados em suspenso, dependendo de investigação adicional.¹⁰

Em outras palavras, Fisher tentava limitar o papel para os valores p de modo que o pesquisador simplesmente jogasse fora o lixo que não merecia uma segunda olhada. No entanto, mesmo essa alegação era duvidosa; de toda maneira, poucos cientistas se interessaram por ela. No começo dos anos 1950, ao descrever a “completa revolução” que o livro-texto de Fisher fizera na pesquisa científica, um proeminente estatístico expressou sua preocupação de que os pesquisadores encarassem “significância” como a essência e a finalidade da pesquisa.¹¹

Seu temor era justificado. Apesar de inúmeras tentativas, os pesquisadores têm se mostrado muito resistentes a abrir mão de suas crenças sobre a significância estatística. Várias vezes buscou-se encarar o assunto. Em 1986, o professor Kenneth Rothman, da Universidade de Massachusetts, editor da prestigiosa *American Journal of Public Health*, disse aos pesquisadores que não

aceitaria resultados baseados unicamente em valores p . A decisão teve efeito dramático: a quantidade de artigos baseados apenas em valores p despencou de mais de 60% para 5%. Todavia, dois anos depois, quando Rothman deixou a editoria, seu veto aos valores p foi abandonado, e os pesquisadores retomaram os velhos hábitos. Outros campos têm tido história semelhante, inclusive a epidemiologia¹² e a economia.¹³

Hoje, apesar de esforços ocasionais de publicações como a *Basp*, pouca coisa mudou. Sociedades acadêmicas têm mostrado notável relutância para lidar com uma questão que “retarda o progresso do conhecimento científico”,¹⁴ enquanto algumas instituições vêm examinando o assunto, porém não tomam qualquer atitude decisiva.¹⁵ Como consequência, as mais importantes revistas de pesquisa continuam a publicar alegações “estatisticamente significativas”, dignas de manchetes que desafiam a credulidade ou as tentativas de resposta. Ao mesmo tempo, novos recrutas das empreitadas científicas aprendem a utilizar os testes de significância – muitas vezes com livros-texto que trazem definições falhas e sem qualquer advertência acerca do significado de tudo isso. A pesquisa mostra que inúmeros estudantes que julgam saber o significado de valores p na realidade não o sabem.¹⁶

O resultado tem sido décadas de perda de tempo, dinheiro e esforço por parte dos pesquisadores – e uma crescente desconfiança nos argumentos científicos entre nós.

Conclusão

Para saber se uma descoberta experimental vale a pena, os cientistas têm como rotina aplicar os chamados testes de significância – apesar de repetidas advertências de que esses métodos são falhos e perigosamente enganosos. O resultado tem sido uma plethora de “avanços” não confiáveis – e uma crescente preocupação com a confiabilidade de argumentos científicos tanto entre pesquisadores quanto em meio ao público leigo.

24. Esquivando-se da espantosa máquina de bobagens

DE TODAS AS CIÊNCIAS, em geral, a física é encarada como a mais dura – e não só no sentido de ser intelectualmente exigente. Suas teorias têm a reputação de sólidas como rocha, baseadas em profunda compreensão dos desígnios do Universo. É discutível quanto essa reputação é merecida; o que não se discute é quanto os físicos têm lançado mão, de modo triunfante, de “big data” (grandes quantidades de dados). Enquanto os pesquisadores das ciências “mais moles”, como a psicologia, muitas vezes precisam se contentar em analisar questionários de algumas dezenas de alunos de faculdade, os físicos gostam de testar suas teorias cósmicas empregando dados pontuais que se contam em bilhões e trilhões. E ninguém faz isso melhor que os físicos experimentais de partículas. Sua meta é desvendar os segredos dos blocos de construção e das forças básicas do cosmo, e as armas escolhidas são máquinas como o Grande Colisor de Hádrons (LHC), com seus 27 quilômetros de comprimento, no Cern, o centro de pesquisa nuclear europeu instalado em Genebra, Suíça. Seu *modus operandi* envolve provocar choques entre centenas de bilhões de partículas subatômicas por segundo, durante horas a fio, e analisar os detritos por meio de sinais significativos de suas pesquisas. A razão de precisarem de tantos dados é que aquilo que estão procurando é em geral incrivelmente raro. Mas, ao longo das décadas, eles se tornaram mestres em encontrar ciscos de ouro científico em montanhas de escória aleatória – e ganham Prêmios Nobel para prová-lo.

Em dezembro de 2011, a equipe do Cern ganhou as manchetes por descobrir a longamente procurada partícula de Higgs, uma fração-chave das teorias unificadoras de todas as forças e partículas da natureza. Cálculos sugeriam que a partícula revelaria sua fugaz existência talvez 1 vez em 1 bilhão de colisões. Entre estas haveria incontáveis eventos aleatórios falsificando a presença de Higgs. Mesmo assim, depois de conferir o resultado de mais de 100 milhões de

milhões de colisões, a equipe anunciou que descobrira a partícula predita pelos teóricos mais de cinquenta anos antes.

A descoberta da partícula de Higgs foi um triunfo difícil, que se assentou sobre a experiência às vezes amarga das peças que a aleatoriedade pode pregar nos incautos – e da inadequação dos métodos convencionalmente usados pelos cientistas para lidar com elas. Tivesse a equipe do Cern seguido a tradição dos pesquisadores em outros campos e declarado sua descoberta usando os métodos padronizados de testes de significância, o anúncio de 2011 teria sido recebido com um ceticismo de revirar os olhos, porque seria apenas mais uma das reivindicações de se ter encontrado Higgs que remontam a meados da década de 1980. Felizmente – e em absoluto contraste com outras áreas da ciência – os pesquisadores em física de partículas há muito vêm insistindo em padrões de evidência muito mais rigorosos antes de ir a público com supostas “descobertas”.

Decerto ninguém no Cern estava ansioso para repetir o vexame de 1984, ocasião em que o laboratório foi a público alegando ter achado outro componente-chave do cosmo que acabou se revelando mero produto da aleatoriedade. Análises dos dados do acelerador haviam indicado a existência do chamado quark top, com massa cerca de quarenta vezes maior que a do próton.¹ A confiança da equipe parecia justificada, de vez que a evidência ultrapassava com folga o padrão adotado em outras áreas da ciência para declarar o resultado “estatisticamente significativo” – implicando a sugestão de que não havia possibilidade de ser um ruído aleatório. No entanto, à medida que surgiam novas evidências, a descoberta mostrou que era apenas isto: um ruído aleatório. Outras descobertas reivindicadas pelo Cern e por um laboratório rival naquele mesmo ano seguiram o mesmo caminho.² O desastre ressaltou as suspeitas havia muito alimentadas pelos físicos de partículas acerca da confiabilidade da significância estatística como medida de evidência. Uma década mais tarde, uma equipe rival nos Estados Unidos disse que encontrara evidência do quark top, mas dessa vez com base num padrão bem mais elevado. A alegação depois foi confirmada inúmeras vezes – assim como a natureza equivocada da “descoberta” do Cern: o

verdadeiro quark top tem massa cerca de quatro vezes maior que a estimada em 1984.

Ao mesmo tempo que os físicos de partículas são conhecidos pelo emprego de máquinas gigantes como o Grande Colisor de Hádrons, eles devem muito do seu sucesso ao ceticismo em relação à espantosa máquina de bobagens que se apresenta sob a roupagem dos “testes de significância”, em cujos resultados os cientistas de outros campos em geral se apoiam. Durante décadas os físicos de partículas testemunharam o irritante sumiço de inúmeras descobertas que haviam passado pelo teste concebido por Ronald Fisher em meados dos anos 1920: resultados com valor p inferior a 5% podiam ser considerados “significativos”. Como vimos no capítulo anterior, os pesquisadores, de hábito, cometem o erro de presumir que isso significa que as chances de o resultado ser pura casualidade também são inferiores a 5%. Alimentada com essa premissa, a espantosa máquina de bobagens transforma casualidades insignificantes em “descobertas” cuja real natureza se torna visível apenas quando alguém tenta confirmá-las. Os físicos de partículas tentaram eliminar os piores excessos da máquina alimentando-a com níveis de significância mais expressivos, geralmente no pacote das chamadas unidades sigma, forma mais elegante e intuitiva de expressar a mesma coisa que os valores p .³ O padrão $p = 5\%$ de Fisher para declarar um resultado “significativo” agora virava um “resultado 2 sigma”, com valores mais elevados de sigma indicando níveis de significância mais altos. Com o passar do tempo, os físicos notaram que até achados 3 e 4 sigma – correspondendo a valores p muito mais “significativos” de 0,3% e 0,006% – também tinham o hábito de sumir diante de novos dados. Em meados dos anos 1990, a principal publicação na área da física declarou que 5 sigma era o nível de significância mínimo aceitável para haver uma reivindicação de descoberta. Pelos padrões da ciência convencional, essa é uma exigência de tirar o fôlego, correspondente a um valor p quase 80 mil vezes mais “significativo” que o nível de 5% de Fisher, comumente usado.

Ainda assim, os físicos de partículas são cautelosos em relação ao que emerge da espantosa máquina de bobagens se ela for alimentada com algum

valor menor. Na comunidade, esse ceticismo é resumido numa regra prática: “Metade de todos os resultados 3 sigma estão errados.”⁴ Essa é uma observação intrigante, que dá indício da fonte de problemas criados ao se confiar na máquina. Se testes de significância representassem o que tantos pesquisadores pensam que sim, então a teoria subjacente aos valores sigma significaria que resultados 3 sigma são casualidades sem significado em média em apenas 1 a cada 370 casos. Todavia, segundo a regra prática, a verdadeira taxa é mais próxima de 1 em 2. Claro que casualidades aleatórias não são a única razão de os experimentos se revelarem não confiáveis. Erros simples também são capazes de minar as pretensas descobertas.

Em 2011 surgiram relatos de partículas chamadas neutrinos viajando mais depressa que a luz. Os dados ultrapassavam o nível de “descoberta” 5 sigma, de modo que o achado provavelmente não era fortuito – e na verdade não era: ele era resultado de equipamento com defeito. Não obstante, a gigantesca discrepância entre o que os pesquisadores *pensam* que a espantosa máquina de bobagens está dizendo e aquilo que eles realmente *obtem* sugere haver algo de seriamente errado na compreensão que têm da máquina. Como vimos no capítulo anterior, de fato há: eles esperam que a máquina realize milagres – a saber, que pegue dados brutos, calcule probabilidades como

$\Pr(\text{observar pelo menos tantos indícios de Higgs, presumindo que sejam por acaso}),$

e aí eles trocam tudo de lugar, na esperança de que esse mesmo número seja a resposta para a questão-chave:

$\Pr(\text{os indícios sejam meramente por acaso, de acordo com quantos indícios observamos})$

O teorema de Bayes nos diz que essas inversões são uma manobra muito arriscada, a menos que tenhamos outra informação – em particular, as probabilidades a priori para aquilo que estamos investigando. Uma vez dadas essas probabilidades, pode-se obter a resposta para a questão-chave que a espantosa máquina de bobagens parece prover, mas simplesmente não consegue.

Contudo, Bayes também pode nos contar o tamanho do erro que cometemos ao confiar na máquina, e os resultados dizem muito.

Peguemos o erro mais comum referente ao valor p : inverter um resultado 2 sigma “estatisticamente significativo” (equivalente ao valor p padrão de Fisher de 5%) e presumir que ele signifique as chances de que nosso resultado seja uma casualidade também é de apenas 5%. Bayes nos diz que só podemos fazer isso se tivermos algum conhecimento a priori do risco de o resultado ser uma casualidade. Como sempre, também confirma a noção de senso comum de que, quanto menos convincente a evidência, mais precisamos estar convencidos de antemão de que os nossos resultados não são casualidade. Pondo a matemática para funcionar,⁵ um fato chocante emerge. Descobre-se que só temos justificativa para interpretar o clássico resultado “ p menor que 5%” como risco menor que 5% de casualidade se *já* estivermos 90% certos de que o acaso não pode ser a explicação. Em outras palavras, a evidência do resultado “significativo” prototípico é tão frágil que não acrescenta virtualmente nada ao nível de crença já existente.

Na realidade, não são apenas os físicos que se tornaram céticos em relação a argumentos baseados em valores p próximos do tradicional ponto de corte de 5% para a significância estatística. Experiências amargas ensinaram aos pesquisadores em muitos campos que o critério de Fisher para a significância simplesmente não é bom o suficiente. Isso fez com que muitos deles atacassem o problema da mesma maneira que os físicos, exigindo evidência mais expressiva – p menor que 0,1%, ou pelo menos 3 sigma – antes de levar a sério novos resultados. Bayes confirma que isso ajuda – mas não muito. Embora pareça 50 vezes mais expressivo que o padrão de Fisher, mesmo esse nível de evidência ainda exige que *já* pensemos que não há mais que um risco de 30% de que o acaso não seja a explicação, antes de levá-la a sério – no sentido de julgar que a evidência nova força o risco para algo abaixo de 5%, como parece implicar o valor p . A verdade é que os pesquisadores na maioria das disciplinas raramente obtêm qualquer coisa próxima a esse nível de evidência.

A boa notícia é que Bayes pode fazer mais do que apenas criar furos na espantosa máquina de bobagens. Ele nos oferece algumas regras práticas para dar sentido àquilo que sai da máquina. Para sermos justos, a máquina de Fisher ao menos tenta dar aos que a alimentam aquilo que eles querem. Em particular, o ponto de corte de 5% para significância especificado pelo seu ilustre inventor mostrou-se popular entre os pesquisadores. Então, vamos pegar o nível de 5% e construir uma versão bayesiana da máquina em torno dele, de modo que signifique o que parece significar: a evidência implica apenas 5% de risco de o resultado ser ditado pelo acaso. Claro que a máquina bayesiana precisará ser alimentada com dados, mas também necessitará do nosso nível de crença a priori – o ingrediente-chave ausente na máquina de Fisher.

NÍVEL DE EVIDÊNCIA (VALOR P)	ÁREAS TÍPICAS EM QUE TAIS NÍVEIS DE EVIDÊNCIA APARECEM	QUANTO VOCÊ JÁ PRECISA ESTAR CONVENCIDO PARA ACHAR ESSE NÍVEL DE EVIDÊNCIA EXPRESSIVO
10%	Economia, sociologia, tópicos “controversos” saúde/ambiente/questões de risco	95% Somente aqueles já convencidos ficarão impressionados.
5%	Quase onipresente; prevalece em especial em ciências médicas, sociais e comportamentais	90% Impressiona apenas se você julga muito improvável que seja por acaso.
1%	Ciências médicas, genética, ciências ambientais	75% Impressiona apenas se você tem muita certeza de que o resultado não pode ser por acaso
0,3%	Estudos de laboratório nas ciências “duras”; alegações preliminares (“3 sigma”) em física de partículas	50% Capaz de impressionar agnósticos de mente aberta.
0,1%	Genética, estudos epidemiológicos	30% Impressiona a todos, exceto céticos, de moderados a severos.
0,00006%	Reivindicações de descoberta em física de partículas e de altas energias	0,1% Muito provável de impressionar a todos, exceto seus rivais.

O que chamaremos de aparelho de inferência bayesiana é, como a máquina de Fisher, uma fórmula,⁶ e ela conduz às seguintes regras práticas. Em cada caso, ela nos dá uma indicação aproximada do nível *mínimo* necessário de crença a priori de o resultado não ser casualidade para que levemos a sério os vários níveis de evidência. Aqui, “levar a sério” significa que a evidência atende ao reverenciado padrão de não mais de 5% de risco de acaso. A tabela também inclui áreas temáticas em que tais níveis de evidência são geralmente declarados como algo ao menos sugestivo, quando não “significativos” ou mesmo convincentes.

A coisa mais admirável em relação aos resultados do aparelho de inferência bayesiana é simplesmente quão tênue a maior parte da suposta evidência “significativa” se revela. Como mostra a tabela, essa evidência caracteristicamente exige que *já* estejamos meio seguros de que os achados não são casualidade antes de termos justificativa para levá-los a sério. E isso, lembre-se, significa “a sério” só no sentido de acreditar que há uma chance de 5% de que sejam por acaso. Se aumentarmos a altura do sarrafo – exigindo, digamos, apenas 1 chance em 100 de estarmos iludidos por uma casualidade –, o nível de exigência necessário vai para as alturas.

Talvez o resultado mais revelador é que se faz necessário um valor p bastante rigoroso (e incomum) de 0,3% antes que mesmo um cético de mente aberta possa se convencer de que o acaso está descartado. Quem for mais cético que isso deve exigir evidência ainda mais expressiva antes de ter confiança suficiente para eliminar a possibilidade do acaso como explicação.

Como sempre, não se deve esquecer que o acaso não é a única razão dos enganos nas descobertas. De fato, pesquisas com supostos valores p surpreendentemente baixos – e, portanto, sigma e níveis de significância surpreendentemente altos – têm reputação de apresentar evidência bastante forte de fenômenos conhecidos como EAL, “Erro em Algum Lugar”.^k A ciência é uma empreitada humana, então sempre refletirá as fraquezas humanas. O aparelho de

inferência bayesiana não pode consertar tudo, mas nos poupa da asneira de seguir pesquisadores proclamando como “significativos” resultados que estão, segundo qualquer definição razoável do termo, muito longe disso.

Conclusão

Muitos “avanços” científicos dignos de manchetes baseiam-se em achados “estatisticamente significativos”. O teorema de Bayes nos leva a regras práticas simples que dão sentido a essas alegações – e muitíssimas delas estão baseadas em evidências tão fracas que devem impressionar apenas aqueles que já “creem verdadeiramente”.

^k Em inglês, ESP, *Error Some Place*. Trata-se de uma brincadeira com Percepção Extrassensorial – em inglês, também ESP, *Extra-Sensorial Perception*, mencionando que o primeiro aparece com frequência no contexto do segundo, e não por mera casualidade. (N.T.)

25. Use aquilo que você já sabe

SE VOCÊ QUER DAR sentido a novas descobertas, o aparelho de inferência bayesiana dá respostas diretas a perguntas diretas – o que é mais do que se pode dizer da espantosa máquina de bobagens e seus valores p . Então, por que ainda há gente usando aquilo que um eminente pesquisador memoravelmente descreveu como “seguramente o procedimento mal orientado mais persistente já instituído no treinamento de rotina dos estudantes de ciência”?¹ Um dos motivos logo fica evidente para quem folhear livros-texto sobre métodos bayesianos. A maioria está carregada de matemática pesada, com pouco interesse aparente para lidar com trivialidades como “Meus achados são só casualidade, ou o quê?”. Isso porque, enquanto Bayes dá respostas diretas a essas perguntas, chegar às respostas pode envolver matemática tão complicada que só é possível com o auxílio dos computadores.² Por muitos anos, essa foi a principal barreira para aqueles que queriam abandonar a máquina de bobagens, mas hoje foi superada, com pacotes de programas padronizados acessíveis para fazer o trabalho pesado.

Mesmo agora, muitos potenciais usuários do teorema de Bayes ficam intimidados pelo secular “problema dos a priori”. Como chegamos ao nível de crença inicial, mesmo antes de termos visto os dados – e será que isso não permite que a subjetividade se infiltre na tarefa científica? Pelo menos, é dessa maneira que em geral se fala do assunto. Mas qual é realmente o tamanho do “problema”? Será que a capacidade de levar em conta aquilo que já sabemos não é uma vantagem? O fato é que, após décadas de pesquisa em muitos campos, temos algumas sacadas bastante boas acerca de muitas coisas, e os métodos bayesianos nos permitem lançar mão disso e contextualizar novos resultados.

O problema é que todas essas sacadas passadas às vezes tiram o brilho das manchetes que anunciam “avanços importantes” e “curas milagrosas” – e ninguém gosta de ser desmancha-prazeres. Basta perguntar a Allen Roses,

executivo sênior do laboratório farmacêutico GlaxoSmithKline (GSK), que em dezembro de 2003 viu-se ocupando as manchetes dos noticiários depois de admitir que, apesar de gastar bilhões na busca de novas terapias, a vasta maioria das drogas não funciona para a maior parte das pessoas.³ Como ressaltou o repórter que deu a notícia, isso não era novidade para os envolvidos na busca de novos tratamentos. Havia muito se sabia que, apesar de todo o alarde sobre as maravilhas da medicina moderna, “curas milagrosas” são poucas e espaçadas, e qualquer alegação em contrário precisa ser encarada com desconfiança.

Apesar disso, ao decidir se devem aprovar ou não alguma nova terapia, os responsáveis pela regulamentação ainda depositam sua confiança nas técnicas de testagem de significância – que não oferecem meios de levar explicitamente em conta experiências passadas. Em contraste, o aparelho de inferência bayesiana aceita de bom grado tanto os dados de estudos quanto os conhecimentos de pesquisas passadas antes de dar uma resposta. E se o argumento alegado voa longe diante da experiência passada, isso pode fazer soar o alarme de uma potencial decepção pela frente.

Em setembro de 1992, pesquisadores médicos na Escócia ganharam os noticiários com os resultados do estudo de uma droga chamada anistreplase. Como trombolítico – “solvente de coágulos” –, a droga pertencia a uma família que já vinha transformando as perspectivas de sobrevivência de pacientes com ataques cardíacos, que recebiam a substância assim que chegavam ao hospital. Dado o benefício da rapidez, porém, parecia inteiramente plausível que a droga salvasse ainda mais vidas se administrada por um médico antes que o paciente chegasse ao hospital. O Great (Grampian Region Early Anistreplase Trial, ou Experimento de Anistreplase Precoce da Região de Grampian) foi montado para descobrir isso, e os resultados que apresentou foram drásticos: as taxas de mortalidade entre vítimas de ataques cardíacos que recebiam a droga antes de chegar ao hospital eram a metade das taxas de pacientes que recebiam a droga no hospital. Considerando como são comuns os ataques cardíacos, esse parecia um avanço importante. Mas os especialistas ficaram reticentes. Afirmaram que, embora alguns benefícios fizessem perfeito sentido, um aprimoramento tão

grande passava muito longe da experiência anterior. Mesmo assim, pelos padrões usuais de se aferir evidência, os achados do Great passavam na inspeção: vinham de pesquisadores respeitados e eram estatisticamente significativos, com um valor p de 4% – dentro do respeitável limite de 5%.

Nos anos seguintes, outras equipes se dispuseram a replicar o avanço, e em 2000 foi publicada uma revisão de toda a evidência, baseada em mais de 6 mil pacientes – 20 vezes o tamanho do estudo Great. A boa notícia era que a técnica de fato parecia oferecer algum benefício; a notícia não tão boa era que, em geral, ela parecia produzir redução no risco de morte de apenas 17% – algo que ainda valia a pena, mas era consideravelmente menos efetivo do que sugerira o estudo original. Em suma, o estudo Great parecia outro caso de avanço que foi sumindo – como sempre, não houve escassez de explicações potenciais, no mínimo mencionando o tamanho relativamente pequeno do estudo inicial. Mas uma explicação se destacou: a que predizia não só que os achados originais iriam se reduzir, mas também em quanto.

Pouco depois de o estudo ser publicado, dois estatísticos britânicos da área de medicina, Stuart Pocock e David Spiegelhalter, escreveram uma breve carta ao *BMJ* argumentando que a redução da taxa de mortalidade pela metade precisava ser *contextualizada*.⁴ Contudo, em vez de recorrer às habituais generalizações vagas, propuseram-se a fazê-lo em detalhes quantitativos, usando o teorema de Bayes.

Em poucas palavras, eles argumentaram que o novo estudo não devia ser visto como um resultado isolado, e sua confiabilidade julgada somente com base nos testes de significância. Em vez disso, alertaram que ele constituía peso de evidência novo, que podia ser combinado com conhecimentos anteriores acerca de drogas trombolíticas e o impacto provável sobre a taxa de mortalidade. Pocock e Spiegelhalter captaram esse conhecimento a priori no chamado “intervalo de credibilidade a priori” – isto é, uma gama de valores entre os quais o risco real de morte tinha probabilidade de se localizar, à luz do conhecimento corrente (ver Box seguinte).

COMO BAYES MOSTROU QUE O GREAT AFINAL NÃO ERA TÃO GRANDE ASSIM

Para fornecer um sumário simples de seus achados, os pesquisadores muitas vezes usam os chamados intervalos de confiança (ICs), que dão uma “cifra principal” e um intervalo para mais ou para menos, refletindo o efeito do acaso. Assim, para o estudo acerca do trombolítico do Great, a equipe resumiu os achados com um IC de 95% para o risco relativo de morte entre aqueles que recebiam o tratamento, em comparação com os que não recebiam, de 0,47 (0,23 para 0,97). Como nenhum benefício relativo daria o valor de 1,0, parece que o tratamento produzia um corte de 53% ($= 100 - 47\%$) no risco de morte, com uma chance de 95% de o benefício chegar até 77%, ou ter um mínimo de 3%. O padrão de 95% é usado por analogia com o padrão de 5% do valor p . Todavia, em relação aos valores p , a interpretação correta de um IC de 95% é ao mesmo tempo técnica e não soluciona a pergunta que queremos responder – necessita-se do teorema de Bayes para deixar as coisas mais claras e relevantes. Em poucas palavras, os ICs padronizados dão apenas 95% de “confiança” de incluir o valor real se presumimos ignorância prévia absoluta de qual poderia ser o valor real, e também assumirmos que apenas o acaso pode minar a descoberta – duas limitações bem expressivas.⁵

Apesar de ainda serem um tanto enganosos, os ICs de 95% decerto são melhores que os valores p , porque contêm mais informação. Se o intervalo exclui valores correspondentes a efeito nenhum – no caso do Great, isso significa o valor de 1,0 –, então o resultado é “estatisticamente significativo”. Como vimos, isso não quer dizer muita coisa. Muito mais significativa, porém, é a *amplitude* do IC – ou seja, a diferença entre os limites superior e inferior. Amostras pequenas são mais vulneráveis aos efeitos do acaso, e se revelam em ICs de grande amplitude. Em termos bayesianos, implicam baixo peso de evidência, e os resultados do Great eram um caso desse tipo. Quando Pocock e Spiegelhalter usaram Bayes para combinar o frágil peso de evidência do estudo com os resultados de duas pesquisas muito maiores, indicando efeitos menos dramáticos, a cifra principal de corte de 53% nas mortes encolheu para 25% – o que, anos depois, acabou se revelando mais realista.

Quando eles combinaram os antigos dados com os achados da nova pesquisa, descobriram que a real eficácia do trombolítico para salvar vidas estava mais provavelmente em torno de 25% – ainda valendo a pena, porém muito menos que o sugerido pelo estudo Great. Os autores lutaram para ter seus resultados divulgados, porém, quando a revisão da evidência foi publicada, sete anos depois, indicando uma redução de 17%, reivindicaram a predição bayesiana.⁶ Essa foi uma demonstração impressionante da importância de se levar em conta a experiência passada e a *plausibilidade* na ocasião de dar sentido aos novos

achados. Mais importante ainda, ao publicar sua predição de uma potencial decepção anos antes dos resultados revistos, Pocock e Spiegelhalter não podiam ser acusados de ter se beneficiado de conclusões retroativas.

Todavia, ao mesmo tempo, eles haviam enfatizado algumas questões importantes sobre o uso de Bayes em assuntos de vida ou morte. Seus cálculos não haviam provado que Bayes realmente permite a todo mundo chegar às suas próprias conclusões, escolhidas a dedo? Suponha, por exemplo, que eles fossem rivais dos primeiros pesquisadores, determinados a exterminar a pesquisa sobre trombolíticos. O que os impediria de selecionar cuidadosamente evidência a priori e fazê-la passar pelo aparelho bayesiano até os resultados do experimento anteriores parecerem estúpidos? Se fossem fãs do tratamento, ou estivessem na folha de pagamento dos fabricantes do trombolítico, com a mesma facilidade poderiam ter desviado os resultados no sentido oposto.

Essas críticas teriam mais peso não fosse o fato de que os pesquisadores sempre desprezaram ou aceitaram as novas descobertas com base em suas percepções ou seus preconceitos – venalidades – pessoais. Os intervalos de almoço em institutos de pesquisa em geral são animados por discussões sobre os novos resultados que ocupam as manchetes dos noticiários, com farto emprego de frases como “Bem, eu ainda não acredito” ou “Você tem de admitir que realmente faz algum sentido”. O uso de testes de significância nada faz para excluir essas práticas descaradamente subjetivas. Isso porque todo pesquisador sabe por experiência própria que, não importa quão expressivo seja o valor p , se o resultado “não cheira bem”, o ceticismo ainda se justifica. O que os testes de significância evitam, porém, é qualquer esperança de colocar isso numa base transparente e *quantitativa*.

Os céticos e crédulos podem se safar com justificativas vagas de menosprezo, e não só durante o almoço: ler as seções de “Debates” de artigos em revistas científicas prestigiosas é ser exposto a uma subjetividade sem limites, travestida em conhecimento especializado. A principal conquista de Pocock e Spiegelhalter naquela breve carta ao *BMJ* foi mostrar que não precisa ser assim. O teorema de Bayes coloca o processo de contextualização dos novos

resultados sobre um alicerce matemático sólido. Obviamente é possível escolher que evidência a priori combinar com os novos achados. A diferença crucial é que o teorema de Bayes obriga céticos e crédulos, em igual medida, a declarar *explicitamente* que evidência a priori estão introduzindo em sua avaliação.

A ideia de macular achados cristalinos com resultados a priori possivelmente falhos ainda pode parecer um risco, mas o aparelho de inferência bayesiana leva isso em conta. Sua mecânica subjacente assegura que, à medida que os dados vão se acumulando, essas crenças a priori tornam-se cada vez menos importantes. A menos que ele seja alimentado com crenças a priori muito, muito esquisitas, tanto céticos quanto crédulos serão levados à mesma conclusão – que não pode ser alcançada por nenhuma discussão durante o almoço.

Conclusão

Avaliar a plausibilidade de novos achados significa colocá-los no contexto daquilo que já sabemos. Com muita frequência, o resultado é só um pouco mais científico do que “Isso soa razoável”. O teorema de Bayes nos dá um meio robusto, transparente e quantitativo de aferir a plausibilidade de novos achados.

26. Desculpe, professor, mas não engulo essa

O MÉTODO CIENTÍFICO TEM muitas conquistas impressionantes em seu crédito. Observatórios em órbita mostraram que o Universo teve início num big bang, cerca de 14 bilhões de anos atrás. Experimentos clínicos nos deram tratamentos efetivos para uma miríade de doenças fatais. E exames de imagem cerebrais de homens assistindo à pornografia mostram que seu cérebro encolhe.¹

Difícilmente se passa uma semana sem que a mídia relate alguma afirmação mais ou menos bizarra baseada em pesquisas publicadas por cientistas reais em periódicos científicos sérios. Tal é a sua presença – e aparente credibilidade – que em 2007 o Serviço Nacional de Saúde do Reino Unido criou um site chamado Behind the Headlines (Por trás das manchetes), onde especialistas analisam essas afirmações e as contextualizam. Inusitadamente, o site não tem a premissa a priori de que todos os jornalistas são mercenários sensacionalistas indignos de confiança, nem que todos os pesquisadores são brilhantes buscadores da verdade. Em vez disso, ele se atém a explicar o que está sendo afirmado e em que medida a alegação se justifica. Numa quantidade enorme de estudos, a resposta é: muito pouco, quase nada. De estudos apontando o efeito milagroso ou fatal de comer ovos a pesquisas sugerindo a ideia do “gaydar” – ou “radar gay” –, que permite às pessoas dizer se os outros são homossexuais,² muitos dos estudos ganham as manchetes porque tratam de questões nunca abordadas. E virtualmente todas elas chegam às suas conclusões via o ritual padronizado de alimentar com dados brutos a espantosa máquina de bobagens.

Mas como Bayes poderia ajudar nesses casos? Afinal, para chegar a funcionar, o aparelho de inferência bayesiana necessita não só de dados brutos, mas também de conhecimentos a priori – e de onde viriam eles quando ninguém fez nada parecido no passado? Esse é um desafio que se torna ainda mais difícil porque os estudos surgidos do nada em geral são pequenos. Como consequência,

não carregam uma porção grande de peso de evidência, e o que existe poderia ser tragado por um a priori mal escolhido.

Estamos nos defrontando de novo com o secular problema dos a priori, e desta vez ele parece especialmente sério. Uma saída é desfraldar a bandeira branca e transformar nosso aparelho na máquina de bobagens – alimentando-a com um a priori “vago” ou “não informativo”. Isso significa assumir que todos os resultados – não importa quão tolos – são igualmente prováveis. Uma resposta menos abjeta seria aceitar que não temos nenhuma evidência a priori para usar, e, em vez disso, buscamos informação em fontes mais genéricas porém menos precisas, ou, como são frequentemente chamadas, nos “peritos”. Isso envolve um processo conhecido como elicitación a priori, que, na sua forma mais simples, inclui fazer os peritos darem palpites chutados sobre intervalos possíveis dentro dos quais eles esperam que se encontre o resultado real. Por exemplo, eles seriam solicitados a declarar um tamanho “mais provável” do efeito, com uma estimativa do nível plausível mais elevado. Isso pode ser combinado de modo a produzir uma “distribuição a priori especializada” geral, alimentada no aparelho para contextualizar o resultado do estudo. Entretanto, trata-se de um processo que tem seus claros perigos. Os peritos podem produzir, e produzem, palpites chutados superimprecisos,³ e estes irão afetar seriamente a interpretação de estudos pequenos. Em todo caso, e se não concordarmos com os peritos? E se mais tarde se provar que eles estavam errados? Como se desconsidera sua influência na interpretação do estudo?

Felizmente, há um botão escondido no reluzente exterior do aparelho de inferência bayesiana que até muitos veteranos em seu uso não chegaram a notar. Ele nos permite não alimentar o aparelho com conhecimentos a priori irremediavelmente vagos ou fornecidos enganosamente por “peritos”, e chegar à nossa própria opinião personalizada da evidência. Em resumo, apertar esse botão faz o aparelho funcionar em marcha a ré. Lembremos como ele geralmente opera: começa com conhecimentos a priori, combina-os com os dados brutos que obtivemos e nos diz se a evidência agora é convincente, à luz daquilo que já sabemos. Mas o aparelho funciona igualmente em marcha a ré, isto é, ele

começa com o que consideramos uma conclusão convincente e trabalha de trás para a frente, de modo a revelar o nível de crença a priori necessário para que os dados justifiquem a conclusão. Logo, em vez de insistir – de forma um tanto absurda – em que “ninguém sabe nada” ou – um tanto pretensiosamente – em que apenas os “peritos” devem estabelecer os a priori, apertar o botão do aparelho permite que cada um de nós dê sentido aos dados nos seus próprios termos. O aparelho nos diz que crença a priori devemos ter para que os dados levem a uma conclusão convincente. Tudo que precisamos decidir é: achamos que esse nível de crença a priori é razoável? Podemos julgar que ele é exagerado demais, e nesse caso estamos inteiramente justificados em encarar os novos achados como não convincentes. Se, por outro lado, não temos problema em incluir esse nível na nossa própria crença, estamos igualmente justificados em alegar que a nova pesquisa conseguiu nos convencer. O processo todo é transparente, democrático e quantitativo – e para muitos tipos de estudo, envolve simplesmente alimentar com dois números uma calculadora on-line.⁴

Mesmo andando em marcha a ré, o aparelho mantém toda a sua potência – inclusive sua capacidade de revelar a verdadeira força da evidência. Peguemos o caso do estudo Great sobre ataques cardíacos, no capítulo anterior, com seu impressionante argumento de redução de 50% no risco de morte se o tratamento for logo administrado. O aparelho lida rapidamente com a alardeada descoberta “estatisticamente significativa” de redução da mortalidade pela metade. Em marcha a ré, ele revela que, para considerar o resultado convincente, a pessoa *já* teria de estar certa de que o tratamento precoce produziria *pelo menos* um corte de 90% na mortalidade. Isso porque o peso da evidência do experimento Great é muito frágil – e, conseqüentemente, seus achados não acrescentam muita coisa ao conhecimento a priori de que já se dispunha. De fato, o estudo Great fracassa em nos persuadir até nos seus próprios termos: seu peso de evidência é tão baixo que a porcentagem de 50% é convincente apenas se já houver evidência para um resultado muito mais expressivo. Isso significa que a pesquisa foi perda de tempo e dinheiro, e metade dos pacientes foram colocados em risco sem razão? Absolutamente não: todo o cerne da pesquisa é ampliar os limites do

conhecimento acumulando evidência. O estudo Great foi parte crucial desse processo, e o aparelho de inferência bayesiana tira o máximo – e dá o máximo de sentido – daquilo que os estudos estão nos dizendo. Com toda a certeza, à medida que se realizaram novas pesquisas com essa abordagem de salvar vidas, mais evidência foi acumulada – e o aparelho mostra que o resultado se tornou cada vez mais convincente. Quando a evidência foi revista, com a resultante notícia de corte de 17% no risco de morte, com base numa quantidade vinte vezes maior de pacientes que a do estudo Great, esse número carregava muito mais peso de evidência, e, portanto, um intervalo de confiança muito mais estreito que 95%. Quando o número é colocado no aparelho em marcha a ré, descobrimos que a credibilidade do novo achado não mais exige que *já* acreditemos ser possível um corte de 90% nas mortes. Agora, para levar a nova evidência a sério, só se requer que julguemos plausível um corte de 28% – exigência muito menor. O aparelho mostra que os novos dados são fortes o bastante para fazer o grosso do trabalho pesado, e não precisam de muita ajuda do conhecimento a priori.

O aparelho pode ajudar a dar sentido até à forma mais estarrecedora de evidência: aquela que “surge do nada”, de pesquisas sobre questões totalmente novas. Esses estudos deixam mesmo os peritos tateando por alguma coisa significativa – quanto mais quantitativa – para dizer. Por exemplo, em 2012 uma equipe da Universidade de Miami foi a público com a alegação de que pessoas que consumiam diariamente refrigerantes diet enfrentavam um robusto – e estatisticamente significativo – aumento de risco, de 43%, de sofrer acidentes vasculares, como um derrame.⁵ Dada a popularidade dessas bebidas de “baixo teor calórico”, e o fato de que o estudo envolvia milhares de pessoas, a afirmação ganhou as manchetes mesmo antes de ser oficialmente publicada. Ainda assim, os próprios pesquisadores estavam preocupados com a possibilidade de seus achados terem ido longe demais. Salientaram que, apesar do tamanho geral do estudo, o número que atraiu as manchetes baseava-se num subconjunto de menos de 10% dos participantes. Os pesquisadores clamavam por estudos muito maiores desse achado potencialmente importante.

Entretanto, o que eles não fizeram, nem ninguém mais fez, foi algo além de simplesmente alimentar com dados brutos a espantosa máquina de bobagens. Se tivessem feito, teriam percebido quão frágil era a evidência. Alimentando com esses mesmos dados o aparelho de inferência bayesiana depois de engatar a marcha a ré, ele nos diz que a cifra de 43%, merecedora das manchetes, só é crível se já estivermos convencidos de que a cifra real é de pelo menos 60%. Mas considerando que esse foi o primeiro estudo a fazer tal alegação, de onde poderia vir a crença? Afinal, nem a própria pesquisa alega uma cifra de risco tão drástica. Em outras palavras, a pesquisa carece tanto de peso de evidência que – como o estudo Great – não chega a ser crível *nem nos seus próprios termos*. O aparelho nos avisa que temos aqui uma evidência estatisticamente significativa na sua forma mais frágil, baseada no fato de que, para considerá-la crível, já devemos acreditar num efeito mais impressionante que o encontrado pelo próprio estudo. Claro que ele colaborou com algum peso de evidência, e essa é uma contribuição potencialmente útil para a ciência. Mas é bem mais preliminar do que insinua aquela cifra de risco robusta e seu rótulo de “estatisticamente significativo”.

Como diz a consagrada frase, mais pesquisa é necessária. Nesse meio-tempo, deveríamos ignorar a cobertura da mídia, deixar os cientistas descobrirem mais e, talvez, em vez disso, ponderar o seguinte. Desde a sua invenção, na década de 1920, testes de significância e valores p vêm confundindo os estudantes, enganando pesquisadores e induzindo o resto de nós a enxergar erroneamente “significância” em resultados que são tudo, menos significativos. Ironicamente, inventados como forma delicada de eliminar casualidades óbvias, eles foram transformados na espantosa máquina de bobagens, que alega revelar o que deve ser levado a sério, mas na verdade não pode. Sejam os resultados da mais recente investigação de um tratamento médico amplamente estudado ou uma alegação surgida do nada sobre algo que nunca ninguém estudou antes, para a máquina é tudo a mesma coisa. Ela apenas absorve os dados, ignora todo o resto e pronuncia seu veredito: é “ouro em pó” ou é “lixo”.

Essa abordagem é inimiga do progresso científico. Em todos os níveis, desde a descoberta da expansão do Universo, passando pela identificação do papel genético do DNA, até a demonstração de que prótons contêm quarks, a ciência tem avançado pela acumulação de evidências, e não por meio de simples dicotomias verdadeiro/falso. Os cientistas captam a realidade em tons sutis, e não em preto e branco. E a maneira de fazer isso é combinar diferentes indícios de evidências utilizando métodos bayesianos.

Mesmo agora, com tantas evidências se formando para demonstrar os fracassos da espantosa máquina de bobagens, essas afirmações ainda são capazes de provocar paroxismos de indignação. Contudo, aqueles determinados a manter a fé na máquina se alinham contra o resultado de um programa de pesquisas que começou justamente quando a máquina estava sendo construída. Durante os anos 1920, diversos matemáticos – especialmente Émile Borel na França, Frank Ramsey na Inglaterra e Bruno de Finetti na Itália – começaram a ponderar sobre a questão de como a evidência concreta é transformada nessa coisa nebulosa chamada crença. O trabalho deles revelou as leis que qualquer abordagem racional e confiável deve seguir. São as leis da probabilidade – com o teorema de Bayes no papel-chave de atualizar crença à luz de evidência. Largamente ignorado durante décadas, esse intrigante elo foi explorado por outros que buscavam dar uma base rigorosa à probabilidade.⁶ Nos últimos anos, encontraram-se as raízes fundamentais da ligação entre inferência e teorema de Bayes, e a ligação se mostra não só meramente plausível, mas efetivamente inevitável.⁷

Em suma, não existe mais nenhuma desculpa para manter a fé na espantosa máquina de bobagens. Ela precisa ser levada para um depósito de sucata antes que provoque mais danos ao trabalho científico. No entanto, algumas partes dela poderiam ser poupadas. Não há dúvida de que a máquina tem uma característica muito atraente, que sem dúvida explica sua duradoura popularidade: ela pode ter dado orientação enganosa sobre a “significância” de nova evidência, mas ao menos era uma orientação clara. A boa notícia é que ainda podemos obter isso do aparelho de inferência bayesiana.

O que temos de mandar para a sucata da ciência, porém, é a ideia de um teste simples tipo passar/fracassar. É hora de todos nós – dos consumidores de evidências científicas a seus criadores – adotarmos um enfoque mais matizado da evidência.

Conclusão

O aparelho de inferência bayesiana nos propicia contextualizar novas evidências, permitindo-nos atualizar o que sabemos. Mas também pode nos ajudar a dar sentido a alguma pesquisa surgida do nada em campos nos quais não se conhece praticamente coisa alguma – e identificar quando a evidência é tão frágil que não nos diz virtualmente nada.

27. A assombrosa curva para tudo

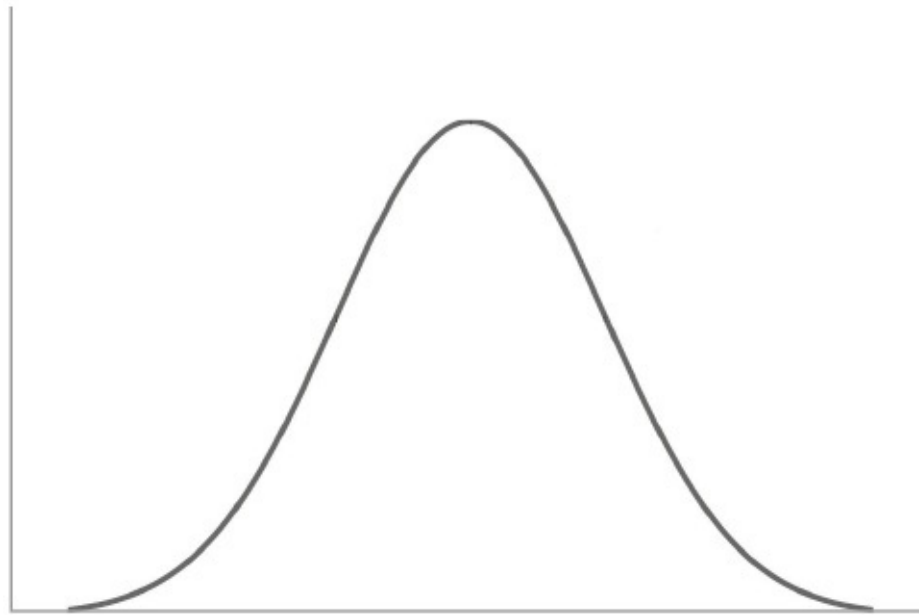
QUANDO OS PRODUTORES DE TV querem que alguém pareça inteligente, fazem questão de garantir que haja algumas prateleiras de livros ao fundo. Quando querem que a pessoa pareça um gênio, substituem as prateleiras de livros por uma lousa coberta de cálculos matemáticos. Há muito eles reconheceram como a mera presença de algumas equações dissipa qualquer dúvida e confere autoridade. Os próprios matemáticos não deixam de ter consciência do poder que sua estranha linguagem exerce sobre aqueles que não a dominam. Segundo a lenda, em 1774, o brilhante matemático suíço Leonhard Euler ganhou um debate público sobre a existência de Deus rabiscando uma fórmula sem sentido num quadro-negro e declarando ser a prova de que Deus existia e exigindo uma resposta. Absolutamente atordoado, seu adversário, iletrado em números, fugiu da sala. Embora a história seja apócrifa,¹ ela fala de uma verdade maior: um dos meios mais efetivos de suprimir divergências é declarar “Há uma álgebra para isso”.

Isso pode ajudar a explicar por que, no fim da década de 1990, administradores seniores de algumas das maiores empresas do mundo ficaram apaixonados pelo seguinte prato de sopa do alfabeto grego:

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

Ficar do lado errado dessa coisinha pode lhe custar o emprego.

Por mais de uma década, funcionários de empresas como Microsoft, General Electric e Conoco podiam se ver, e de fato se viam, demitidos por se colocar do lado errado da fórmula – ou, mais precisamente, da curva que ela descreve, mostrada a seguir:



A bela, sedutora e absolutamente perigosa curva do sino.

É a famosa curva do sino, e por certo tempo os departamentos de recursos humanos estavam convencidos de que ela captava com precisão matemática a performance de seus funcionários, mensurada segundo qualquer escala que se imaginasse: vendas, lucros, “eficácia”, o que fosse. A curva representava graficamente as supostas verdades incorporadas na fórmula. Primeiro, que a maioria dos funcionários tem performance perto da média, e se localizam perto da “corcova” central, com metade da equipe acima da média e a outra metade abaixo. Segundo, uma pequena proporção de funcionários é formada por verdadeiras estrelas, com performance excepcional que os coloca na “cauda” da direita da curva do sino. E terceiro, havia uma proporção correspondente de vagabundos, fracassados e parasitas, todos amontoados na cauda da esquerda. Estes podiam ser identificados, convocados para uma conversa séria ou demitidos. Mas como fazê-lo? Simples: avaliar a performance da equipe numa escala de 1 a 5, certificando-se de que as proporções correspondentes a cada avaliação sigam os ditames da curva do sino. Assim, a maioria deveria tirar um escore médio de 3, enquanto um pouco menos deveria tirar 2 ou 4. Resolvido o caso destes, a gerência podia então focalizar os “atípicos”. Os da “cauda” direita, com avaliação de 5, receberiam gratificações, enquanto suas contrapartes da esquerda seriam chutadas para fora.

Não foi surpresa nenhuma que essa bizarra rotina provocasse considerável ressentimento entre os empregados – e também desconfiança. Muitos sentiam que havia algo não muito correto em relação àquilo que se tornou conhecido como Rank and Yank, algo como “avaliação e descarte”. Alguns deles, descobrindo-se na cauda errada da curva do sino, resolveram levar seus empregadores aos tribunais. Todavia, poucos se sentiram confiantes para atacar a fórmula em si. De modo surpreendente, demorou mais de uma década para que o feitiço matematicamente induzido fosse quebrado. Sim, a fórmula está correta do ponto de vista matemático, e, sim, não há dúvida de que a curva do sino reflete muitas características humanas, tais como altura e QI. Mas ninguém pensou em checar se “performance” era uma delas. Quando checaram, os resultados confirmaram o que muita gente suspeitava: o intervalo está longe de ser simétrico.² Em vez disso, geralmente são apenas alguns poucos os que têm performance muito elevada. A ideia de que *deve* haver uma proporção igual de estrelas e fracassados em todo departamento acaba se revelando – sem nenhuma surpresa – muito mais que estúpida, e uma séria ameaça ao bem-estar corporativo. Forçando as avaliações a se conformar com os ditames da curva do sino, os administradores viam-se obrigados a repreender, digamos, 10% dos funcionários simplesmente porque 80% precisam se encontrar na média, ou perto dela – deixando 20% nas duas “caudas”. No final, a falta de evidência de que essa atitude não conseguia nada a não ser arrasar o moral dos empregados levou muitos antigos defensores a abandonar avaliações com base na curva do sino. A Microsoft e diversas empresas mudaram de atitude, mas inúmeras outras persistem. Algumas podem até ter uma boa causa, porém a chance é de que permaneçam enalhadas numa das armadilhas mais profundas ao lidar com a incerteza: a crença de que praticamente tudo é normal.

Essa poderia parecer uma crença perfeitamente razoável, mas aqui a capitalização é crucial. Pois, como tantos outros termos na teoria da probabilidade e da incerteza, normal tem um significado muito específico, que quase convida ao abuso. Parece implicar algo comum, padronizado ou natural, porém, nesse caso, significa conformidade com os ditames da curva do sino –

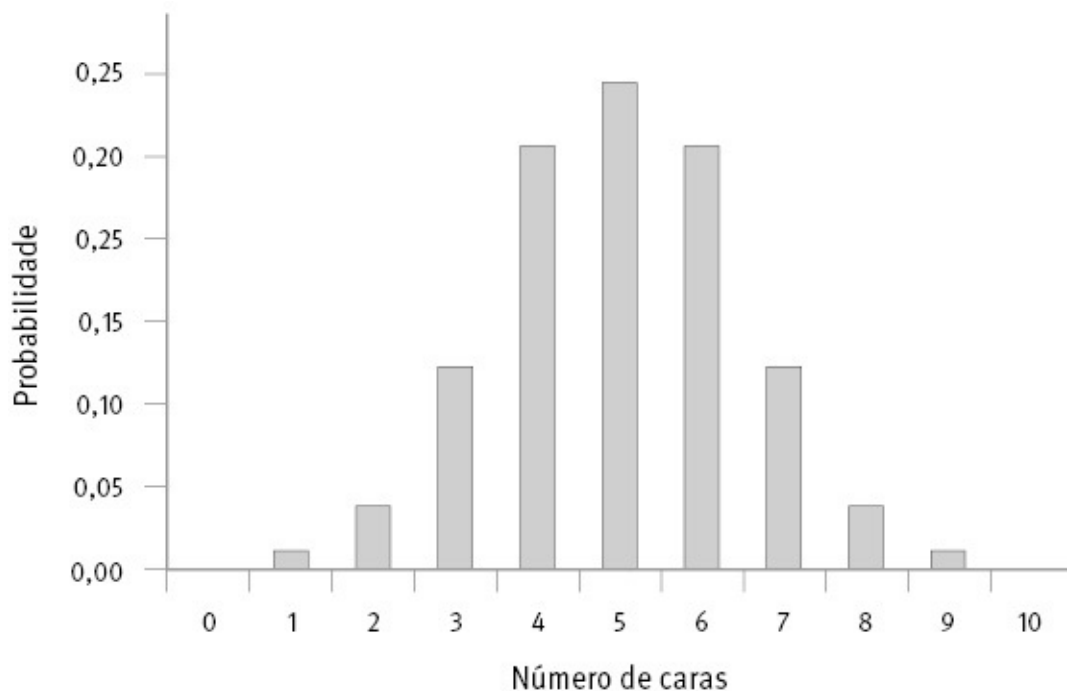
ou, como os matemáticos a chamam, da distribuição normal, cuja fórmula já foi dada. Na verdade, o termo é duplamente inadequado, pois a distribuição normal não só falha com frequência na descrição de fenômenos “normais”, como a fórmula por trás dela é resultado de uma das mais excepcionais descobertas matemáticas já feitas.

Suas raízes se estendem até os próprios primórdios da teoria da probabilidade. Durante o século XVII, os pioneiros do campo – entre eles Pascal, Fermat e Bernoulli – haviam descoberto maneiras de calcular as chances de diferentes combinações de eventos, como, por exemplo, obter três 6 em 10 lançamentos de dado. As respostas emergiram a partir de fórmulas que incluíam tanto as chances de ocorrer o evento individual numa única tentativa quanto o número de maneiras (“permutações”) em que o evento especificado podia aparecer durante os lançamentos. Por exemplo, três 6 podiam aparecer em sequência ou em intervalos aleatórios. No entanto, algo intrigante emergia quando os resultados eram dispostos no papel: à medida que aumentava o número total de tentativas, as chances de obter um número específico de sucessos pareciam seguir uma curva bem distinta.

Essa característica aparecia até nas manifestações mais simples do acaso, como atirar uma moeda. Dado que as chances de tirar cara em qualquer lançamento são 50:50, seria de esperar que o número mais provável de caras fosse a metade do número total de lançamentos. No entanto, ao colocar num gráfico os resultados da fórmula no caso dos lançamentos, cria-se um pico de probabilidade em 5 – o número médio de caras que poderíamos esperar obter. As fórmulas também davam as chances de obter outra quantidade de caras em 10 lançamentos – mostrando inclinações íngremes de cada lado do pico central, refletindo sua menor probabilidade de ocorrer. Em cada extremidade estavam os eventos mais raros de todos: nenhuma cara ou só caras, em 10 lançamentos.

Fazer os cálculos desses gráficos no papel não é para quem tem coração fraco. Até um matemático magistral como Jacob Bernoulli lutou para lidar com qualquer coisa além de pequenos números de tentativas.³ Todavia, sem executar esses cálculos, era difícil descobrir muita coisa sobre as curvas. Necessitava-se

de um tipo de atalho, e em 1733 a solução chegou com uma fórmula não só mais fácil de usar, como também se tornava mais confiável à medida que o número de tentativas aumentava. Essa fórmula fora concebida por Abraham de Moivre (1667-1754), emigrado francês professor e consultor de matemática que vivia em Londres, um dos mais brilhantes matemáticos de sua época. Os talentos de De Moivre na teoria da probabilidade eram tais que até o imperioso gênio Isaac Newton teria recorrido a ele nesses assuntos. Ironicamente, De Moivre também era um pouco azarado, tendo deixado de receber crédito por diversas descobertas – inclusive sua elegante fórmula para probabilidades. Em vez disso, a fórmula ganhou vários títulos, inclusive curva de Gauss, em honra ao grande matemático alemão Carl Gauss (1777-1855), que a descobrira seguindo caminho completamente diferente.



A curva do sino se ergue: as chances de obter diferentes quantidades de caras para 10 lançamentos de uma moeda.

Na época, Gauss lutava com um dos problemas centrais da ciência experimental: extrair informações de dados sujeitos a erro. Ele mostrou que, elaborando três premissas razoáveis sobre como os erros afetam as observações,

podia calcular as chances de o valor real se situar num intervalo específico. Sua fórmula era essencialmente a mesma que a achada por De Moivre, e é aquela que está no começo deste capítulo. Quando posta num gráfico, ela também forma a curva do sino. De Moivre já havia demonstrado que o pico central coincidia com o resultado mais provável de um dado número de eventos aleatórios como lançamentos de moeda, e era algo conveniente para jogadores que quisessem calcular as chances de uma aposta valer a pena. Mas Gauss provara que o pico também representa a média de um conjunto de medições, cada qual sujeita a um erro casual. E isso a tornava imensamente útil para os cientistas que tentavam aferir o provável intervalo dentro do qual o valor real de uma grandeza pode se situar.

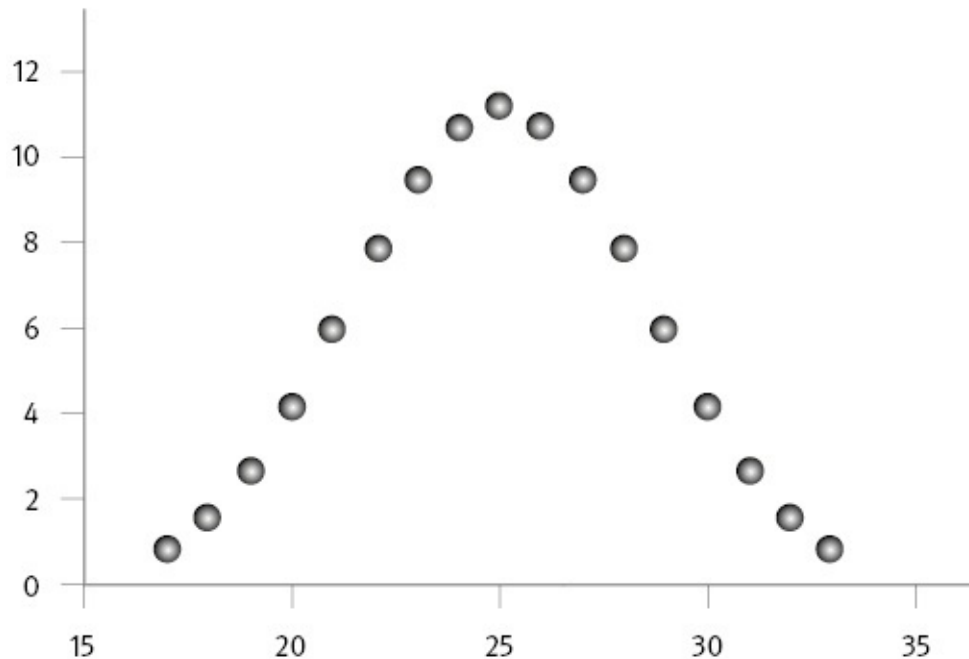
A primeira aparição publicada da fórmula tornou Gauss internacionalmente famoso com apenas 24 anos de idade. Em 1º de janeiro de 1801, um astrônomo italiano fez sensação alegando ter achado um novo planeta no sistema solar, orbitando entre Marte e Júpiter. Infelizmente, antes que alguém pudesse confirmar a descoberta, o objeto perdeu-se sob o brilho intenso do Sol. Sem conhecer sua órbita, havia o risco de o planeta não voltar a ser encontrado durante anos. Gauss aplicou sua fórmula para extrair o máximo valor a partir das observações existentes e – após alguns cálculos assustadoramente difíceis – predisse onde o objeto deveria reaparecer. Dito e feito. Usando os vaticínios de Gauss, os astrônomos “recuperaram” o objeto no mesmo ano. Batizado de Ceres, ele era o maior de um enxame de planetas menores, os chamados asteroides, orbitando o Sol.

Enquanto era saudado pela assombrosa realização, o próprio Gauss guardava dúvidas acerca da base da sua fórmula de erro. Felizmente, ela foi assentada sobre um alicerce sólido graças a outra descoberta, de significado bem maior que a descoberta de Ceres.

A descoberta foi feita por outro titã da matemática aplicada do século XIX, Pierre Simon de Laplace (1749-1827). Já celebrado por achados importantes em cálculo e mecânica celeste, o brilhante polímata francês voltou sua atenção para a probabilidade. Em 1810, revelou algo sobre a curva do sino que mesmo De

Moivre e Gauss tinham deixado escapar. Numa verdadeira façanha mental e matemática, Laplace mostrou que as raízes da curva do sino penetravam bem mais fundo do que se suspeitava, conferindo-lhe enorme importância. Essa era nada menos que uma lei da natureza – lei que, era de esperar, devia estar à espreita numa hoste de fenômenos, incluindo alguns aparentemente destituídos de qualquer causa ou razão. Indícios podem ser encontrados na curiosa ubiquidade de curvas em forma de sino no resultado de eventos aleatórios como cara ou coroa. Apesar de cada lançamento da moeda ser aleatório e completamente independente, quando seu efeito combinado é computado em massa, de algum modo os resultados conspiram para produzir a mesma forma.

Por exemplo, se 100 pessoas forem persuadidas a lançar uma moeda 50 vezes cada e a anotar a quantidade total de caras observadas, cerca de uma dezena de pessoas obterá o total esperado de 25 caras. Cerca de 50 pessoas obterão resultados dentro de um intervalo de mais ou menos 2 em relação a esse valor médio. Mas, além disso, o número de pessoas obtendo totais mais distantes da média começa a decair bem depressa. Mal chegará a uma dezena a quantidade de pessoas a obter resultados situados a mais de 5 em relação à média, enquanto apenas uma ou duas terão azar suficiente para obter menos que 17 ou sorte bastante para observar mais de 33. Marcados os valores num gráfico, o resultado será uma curva em forma de sino mostrando quantas pessoas obtêm vários totais de caras.



Se 100 pessoas lançam uma moeda 50 vezes, só cerca de uma dezena pode esperar obter exatamente 25 caras.

A monumental descoberta de Laplace foi que a mesmíssima curva do sino descreverá *qualquer* fenômeno resultante de um efeito combinado de *qualquer* tipo de influência aleatória agindo independentemente entre si. De forma incrível, não precisamos saber com exatidão quais são essas influências ou como se comportam. Grosso modo, enquanto forem muitos, do mesmo tipo e agindo de forma independente, seu efeito combinado produzirá a curva do sino.⁴

Se você está pelejando para entender as implicações disso, veja-se em ótima companhia: nem o próprio Laplace nem seus contemporâneos entenderam de imediato todo o seu significado. Levou pouco mais de um século para que a descoberta de Laplace adquirisse um título refletindo seu papel-chave na compreensão da incerteza. Por direito, deveria se chamar lei fundamental de influências aleatórias. Na realidade, é conhecida pelo surpreendente e prosaico título de teorema do limite central.

Sua aplicabilidade, porém, é qualquer coisa menos monótona. Considerando que muitos fenômenos poderiam ser razoavelmente pensados como efeito cumulativo de uma miríade de influências aleatórias, seria de esperar que a curva

do sino fosse onipresente. Com toda a certeza, é encontrada em tudo, desde a agitada trajetória das moléculas de um gás, passando por notas de exames de alunos, até o calor que resta do big bang. O exemplo quintessencial é a curva do sino da estatura humana. Considerando que a altura é a soma total dos comprimentos dos vários ossos, cada um resultado de uma miríade de influências, de genes e nutrição até a condição geral de saúde, seria esperável que a curva do sino aparecesse quando a proporção de pessoas com alturas diferentes fosse posta num gráfico em relação a vários intervalos de altura. E *voilà!* É exatamente o que aparece.⁵

O teorema do limite central, no entanto, faz mais do que apenas suprir lastro para argumentos leves. Sua espantosa generalidade oferece uma capacidade quase milagrosa de fazer cortes através da complexidade. Em nenhum lugar isso é mais visível que na pesquisa médica. Para descobrir se uma nova terapia funciona, os clínicos recrutam pacientes e os dividem aleatoriamente em dois grupos, um que vai receber o novo tratamento e outro que receberá uma terapia alternativa. Essa alocação aleatória reduz o risco de que qualquer um dos grupos seja de algum modo anormal, aumentando assim as chances de que os resultados representem o futuro típico do paciente. Obviamente é impossível levar em conta cada detalhe da reação de um paciente, mas o teorema do limite central torna isso desnecessário. Enquanto esses “desconhecidos desconhecidos” afetam cada paciente de forma independente, seu efeito cumulativo será uma curva do sino para cada grupo de pacientes. E se os picos estiverem distanciados o suficiente, será difícil desprezar a diferença como alguma afortunada casualidade.

Os médicos são mais cômicos que a maioria das pessoas da presença do termo “limite” no nome do teorema. Este é um reflexo do fato de que ele vale estritamente apenas no caso de um número infinito de variáveis aleatórias. Na realidade, ele faz um trabalho bastante bom com números relativamente pequenos; mesmo assim, a menos que um experimento clínico envolva pacientes suficientes, há o risco de o verdadeiro impacto da droga ser engolido pelos “desconhecidos desconhecidos”. Para lidar com isso, os clínicos invertem o teorema para estimar aproximadamente quantos pacientes necessitam incluir

para ter uma chance razoável de demonstrar que a terapia de fato funciona – o que é revelado por duas belas e distintas curvas do sino, uma para cada grupo.

O teorema do limite central sem dúvida é uma das ferramentas mais poderosas já entregues pelos matemáticos aos cientistas. Sua pura generalidade é sedutora, e, ao oferecer uma sustentação rigorosa para a curva do sino, deflagrou uma revolução ao aplicar a matemática aos embaralhados fenômenos do mundo real. Mas também se tornou o retrato típico daquilo que pode dar errado se uma ferramenta matemática é mal utilizada e seus “termos e condições”, ignorados. O poder do teorema fez com que ele passasse a ser embutido em técnicas usadas extensivamente em ciência, tecnologia, medicina e negócios. Todavia, pouca gente tem conhecimento de sua presença, muito menos do risco de desprezar os “termos e condições” que governam seu emprego. Como resultado, o teorema de Laplace e sua prole com frequência são forçados a ir longe demais, resultando em achados de pesquisa não confiáveis, descobertas absurdas e um papel central na maior crise financeira dos tempos recentes.

Os sinais de alerta já eram visíveis mesmo quando Laplace ainda tentava entender as implicações de sua descoberta. Em agosto de 1823, um astrônomo belga chamado Adolphe Quetelet (1796-1874) entrou no ilustre Observatório de Paris, em “uma das viagens breves mais famosas na história da ciência”.⁶ Seu objetivo era preparar-se para dirigir um novo observatório em Bruxelas, em particular, compreender a melhor maneira de extrair conhecimento a partir de dados. Quetelet encontrou-se com muitos luminares da ciência, inclusive Laplace. Mas, depois de ver como a curva do sino podia ser usada para descrever erros de observação, começou a refletir sobre aplicações mais empolgantes. Poderia a curva estar ocultamente à espreita em dados sobre características dos seres humanos?

Quetelet imaginou captar todas as qualidades essenciais da humanidade com curvas do sino, e um dia revelar o ser humano prototípico, ou, nas suas palavras, *l’homme moyen* – o “homem médio”. Nos anos que se seguiram, ele começou a publicar evidências para respaldar essa noção. Coletando e reunindo dados sobre uma legião de traços humanos, Quetelet começou a achar curvas do sino em todo

lugar, desde medidas do peito de soldados até a propensão para casar-se ou cometer crimes. Convencido de que tinha descoberto uma “lei” da natureza humana, começou a empregá-la para extrair informações de conjuntos de dados. Lançando mão da curva do sino que captava as alturas dos seres humanos, Quetelet comparou a curva para a população masculina geral da França com aquela de homens recrutados para o exército em 1817. Com as demais variáveis iguais, as curvas deveriam ser iguais, mas não eram: havia uma curiosa “esquisitice” na distribuição próxima à altura-limite para o alistamento. Quetelet acreditou que sua lei havia revelado que cerca de 2% dos homens chamados para o serviço militar haviam evitado o alistamento mentindo sobre a altura.

Não levou muito tempo para que o trabalho de Quetelet com a curva do sino começasse a ser visto como suporte para o emergente conceito de “ciência social”, com todos os tipos de traços humanos encarados como a soma total de influências aleatórias invisíveis. O próprio Quetelet acreditava que a onipresença da curva era uma manifestação da lei dos erros, conforme investigada por Gauss e Laplace. Para ele, o “homem médio” representava a perfeição, e todos os desvios eram resultado de “erros”. No entanto, alguns buscavam uma explicação menos metafísica, e acreditavam que a tinham descoberto no teorema de Laplace. Para eles, a pura onipresença da curva do sino simplesmente refletia a pura onipresença de fenômenos resultantes de influências aleatórias associadas. Laplace, ao que parecia, construía uma ponte entre o mundo platônico da matemática e o atrapalhado mundo da vida real. Quem resistiria a atravessar essa ponte?

Decerto esse não era o caso do polímata vitoriano Francis Galton (1822-1911), que, mais que ninguém, convenceu-se da universalidade da curva do sino.⁷ Em 1877 ele começara a referir-se àquilo que os matemáticos alternadamente chamavam de lei dos erros ou lei de Gauss-Laplace por um nome absolutamente mais potente: lei normal. A implicação era clara: a curva do sino refletia o comportamento típico de fenômenos naturais, o estado habitual das situações, o modo-padrão das coisas. Outros pesquisadores influentes passaram a

fazer o mesmo, entre eles Karl Pearson (1857-1936), um dos fundadores da estatística moderna.

Mas outros estavam preocupados com uma circularidade perigosa subjacente à crença de que a curva do sino era “normal”. Entre eles, o distinto matemático francês Henri Poincaré (1854-1912) e o físico ganhador do Prêmio Nobel Gabriel Lippmann (1845-1921), que comentou sombriamente: “Todo mundo acredita nisso – experimentalistas acreditam que é um teorema matemático, matemáticos acreditam que é um fato empírico.”⁸

Como veremos, sua preocupação acerca dessa garantia mutuamente destrutiva revelou-se presciente demais.

Conclusão

De todas as leis subjacentes ao comportamento dos efeitos do acaso, nenhuma é mais atraente que o teorema do limite central de Laplace, e sua explicação para a aparentemente ubíqua curva do sino. Mas não se deixe cair nas conversas de livros-texto sobre “distribuição normal” – porque normal ela não é.

28. Os perigos de pensar que tudo é normal

AO LONGO DOS SEUS mais de 150 anos de história, o banco de investimentos Goldman Sachs viu de tudo. Booms econômicos, colapsos financeiros, bolhas de ações, recessões globais – fosse qual fosse a crise, ele continuava em sua rota por todas elas. Mas, em agosto de 2007, o banco se chocou contra o equivalente financeiro a uma frota de icebergs, e teve de enfiar mais de US\$ 2 bilhões em dois fundos para impedir o naufrágio. Como principal responsável financeiro do banco, David Viniar deveria ser o vigia atento na ponte de comando. Então, como deixou de ver esses leviatãs? O relato que fez para um repórter nesse dia tornou-se matéria de lenda entre os conhecedores das finanças: “Víamos coisas que estavam a 25 desvios-padrão, vários dias seguidos.” O que, traduzido para linguagem comum, significa: “Tivemos muito azar.”

Ou pelo menos foi o que entenderam os fluentes em “papo quantitativo”. Essa é a linguagem dos analistas quantitativos, pessoas que, como Viniar, usam modelos matemáticos para compreender o risco e a incerteza no mundo financeiro. Esses analistas – conhecidos como *quants* – carregam um bocado de coisas surpreendentes na cabeça, incluindo certos números básicos que lhes permitem dar sentido imediato a dados financeiros novos. Todos eles sabem, por exemplo, que um movimento de mercado de “1 sigma” tem 68% de probabilidade de ocorrer por acaso, e é tão comum que ninguém perde o sono por causa dele. Mas um evento “2 sigma” tem probabilidade de apenas 5%, o que o torna oficialmente um desvio “estatisticamente significativo” da conduta usual. Ainda assim, a coisa acontece. É muito mais difícil ficar otimista com um evento 4 sigma; agora estamos falando de cerca de 1 chance em 16 000 de isso acontecer por acaso. Você poderia passar sua carreira inteira sem viver um dia desses. Contudo, mesmo os *quants* mais experientes não teriam cabeça para enfrentar os eventos 25 sigma de Viniar. São tão bizarros que até as fórmulas

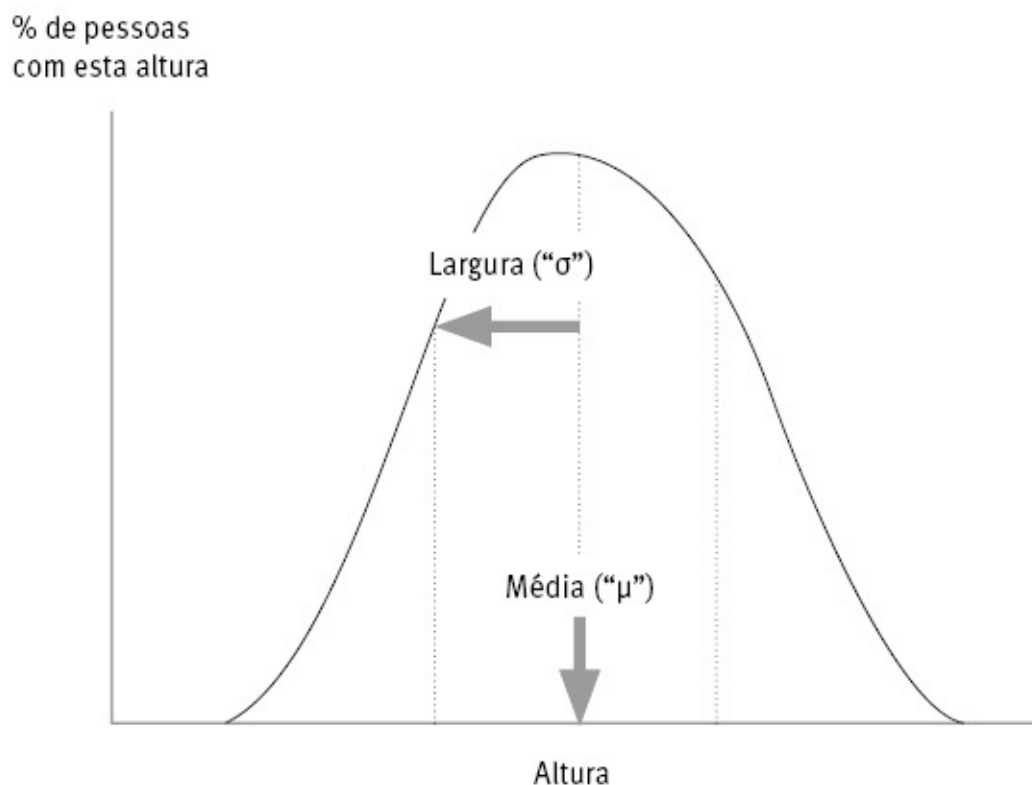
padronizadas para eles caem por terra, e tornam-se necessárias medidas especiais para conseguir que as planilhas exibam as chances de tais eventos, de tão baixas que são.¹ Mas quando finalmente são instigadas a dar uma resposta, é algo verdadeiramente assombroso. Viniar e seus colegas tinham alegado que foram surpreendidos por um evento que deveria ocorrer em média apenas 1 vez em cada 10^{135} anos. Esse é um número para lá de astronômico; é uma escala de tempo inconcebivelmente mais longa que a idade do Universo. E, segundo Viniar, sua empresa estava na extremidade receptora de *vários* desses eventos.

Ainda que não haja motivo para duvidar da cifra 25 sigma de Viniar, a impressionante baixa probabilidade que ela implica é problemática. Claro, eventos raríssimos podem ocorrer e ocorrem o tempo todo. Mas quando vários deles acontecem todos juntos, é de estranhar; será que há algo de errado com a maneira pela qual as chances foram calculadas? Fazer os cálculos exige a chamada distribuição de probabilidade para o evento. Esta aparece em grande quantidade de formas e tamanhos, mas, em finanças, há uma para a qual todos se voltam quase sem pensar: a curva do sino. E por que não? Afinal, ela não é literalmente a distribuição normal?

Preocupações relativas à premissa rotineira de “normalidade” emergiram quase na mesma época em que o conceito começou a se tornar corrente, mais de um século antes da declaração de Viniar. Em 1901, o estatístico pioneiro inglês Karl Pearson examinara algumas das alegações feitas a favor da suposta onipresença da curva do sino, e descobriu evidência menos que convincente. Escreveu: “Eu só posso reconhecer a ocorrência da curva normal ... como um fenômeno muito anormal.” Já nos anos 1920, Pearson lamentava ter ajudado a criar a ilusão de que a curva do sino fosse “normal”, declarando que o termo “tem a desvantagem de levar as pessoas a acreditar que todas as outras distribuições são, em um ou em outro sentido, ‘anormais’”.² Ele recomendou que a premissa de normalidade fosse somente um primeiro palpite nos estudos teóricos. Contudo, esses escrúpulos foram varridos para longe quando a curva do sino se tornou não só o primeiro palpite, mas o único. Ela era simplesmente

elegante demais, a lógica para sua ubiquidade era convincente demais, o encaixe em tantos conjuntos de dados, impressionante demais.

Mas exatamente quanto os dados da vida real se encaixam na curva do sino? O exemplo dos livros-texto é a altura das pessoas, e a receita, muito simples. Primeiro, pegue medidas de um monte de gente e marque as porcentagens daquelas pessoas cujas alturas caem nas várias faixas – com diferenças de 5 milímetros. O resultado é um gráfico de barras cujo contorno forma uma curva do sino bastante regular, mais ou menos parecida com isto:³



A bela, ainda que levemente dentada, curva do sino das alturas humanas.

Também obtemos a demonstração verdadeira do poder da distribuição normal para sintetizar uma enorme massa de dados em apenas dois números. O primeiro é a altura média, representada pela letra grega μ ("mi"). Esse número localiza o pico central da curva do sino em relação ao eixo horizontal. Depois há o desvio-padrão, representado por σ ("sigma"), que descreve a largura da curva do sino. Uma vez encaixando uma curva do sino nos dados, o conhecimento

desses dois números basta para nos dar uma enorme quantidade de informações. Por exemplo, 95% da porcentagem total da curva está quase exatamente num intervalo de mais ou menos 2 sigma da média. Assim, por exemplo, se sabemos que a altura média é de 175 centímetros, e que o desvio-padrão é de 7,5 centímetros, sabemos que 95% das pessoas terão altura entre cerca de 160 e 190 centímetros. Isso, por sua vez, quer dizer que 5% das pessoas se encontram fora desses limites. Como a curva é perfeitamente simétrica, podemos dividi-las exatamente em 2,5% que são mais baixas que 160 centímetros e 2,5% que são mais altas que 190 centímetros. Podemos também inverter esses cálculos e perguntar que porcentagem de pessoas têm altura maior que, digamos, 4 desvios-padrão (“4 sigma”) acima da média. A fórmula para a curva do sino mostra que cerca de 1 em 16 000 se encontra além de 4 sigma da média, logo, por simetria da curva, exatamente metade dessa proporção estará acima. Num país de, digamos, 100 milhões de pessoas com essa distribuição de altura, esperamos encontrar cerca de 3 000 com altura maior que 205 centímetros.

Tudo isso é maravilhoso, e é difícil não se sentir inundado de poder. No entanto, há só um problema – bem visível, se olharmos com mais cuidado para a curva do sino da vida real. Ainda que seja uma curva *em* forma de sino, não é *a* curva do sino. O teorema de Laplace é bastante firme nesse ponto. Para qualquer fenômeno que se presuma seguir seus ditames, o teorema do limite central nos diz que obteremos uma única e bela curva simétrica, com caudas graciosas descendo suavemente de ambos os lados. No entanto, obtivemos uma curva parruda, atarracada, com um pequeno entalhe perto do pico. Então, o que deu errado? Talvez não tenhamos coletado dados suficientes para contrabalançar todas as irregularidades. Isso é possível, mas não ajuda muito, pois nunca estaremos *absolutamente* certos de termos obtido uma curva do sino perfeita, porque a teoria exige, para isso, que haja um número infinito de pontos dados. O que mais pode causar problemas? Picos dentados talvez sejam sinal de que inadvertidamente misturamos duas populações diversas, com diferentes influências aleatórias em ação.

No caso das alturas humanas, podemos ao menos dar um palpite de quais sejam essas “populações diferentes”: homens e mulheres. Decerto, se separarmos os dois gêneros, obteremos curvas do sino de aparência melhor, mas ainda assim menos que perfeitas. Tudo bem, então talvez seja porque não basta apenas dividi-las em duas populações – talvez haja subgrupos dentro de subgrupos. Isso faz sentido: haverá o contexto étnico, o estado nutricional e sei lá mais o quê. Agora deparamos com outro problema: o teorema de Laplace exige que todos esses efeitos aleatórios diferentes ajam de forma independente para nos dar uma curva do sino real. Mas será que isso é plausível? Genes sem dúvida não agem de forma independente uns dos outros, tampouco as influências nutricionais – e argumentar que todos esses fatores exercem apenas efeitos aditivos, como exige o teorema de Laplace, é um triunfo da esperança sobre a experiência. Em suma, espantoso é que o topo de curva tenha um aspecto remotamente redondo e regular, ou que suas encostas laterais sejam simétricas.⁴

Alguns dos primeiros defensores da teoria da “onipresença da curva do sino” reconheceram esses questionamentos. Quetelet tentou separar dados específicos de gênero, e os resultados foram suficientemente bons para seus estudos do *l’homme moyen*, que por definição se encontra no pico da curva. Mas alguns de seus contemporâneos buscaram forçar a barra com a curva do sino, a fim de descobrir o que ela dizia sobre os *extremos*. Isso os afastou da relativa segurança do pico da curva, levando-os para as caudas. Ao fazê-lo, eles deixaram de notar – ou optaram por ignorar – que estavam em perigo cada vez maior de perder contato com a realidade. Observe qualquer coleção de dados reais sobre qualquer coisa, e não importa quantos você tenha, sempre será capaz de encontrar duas classes: os maiores e os menores. Claro que pode haver ainda maiores ou ainda menores em algum lugar por aí, talvez uma enorme quantidade deles. O problema é que você não sabe; a única certeza que você tem é de que, ao coletar seus dados, sempre acabará com dois extremos, nada além deles.

Mas a bela curva teórica de Laplace *jamaís* acaba. Suas caudas continuam a descer suavemente para sempre, só beijando o eixo horizontal no infinito. E isso tem uma implicação surpreendente para quem tenta usar a curva do sino para

imitar a realidade. No caso da altura humana, por exemplo, significa que há uma chance – embora minúscula – de encontrar seres humanos mais altos que o monte Everest, e alguns sem altura nenhuma, ou até mesmo alturas *negativas*. Como as probabilidades de qualquer um desses absurdos são pequenas, é tentador tratá-las apenas como mais uma esquisitice, como o pico dentado. Ainda assim, mesmo com Quetelet e seus contemporâneos enxergando curvas do sino por toda parte, havia uma real prova viva de que não se podia confiar na curva do sino nos extremos. E essa prova assumiu a admirável forma de Bud Rogan, o Homem Impossível.

Nascido no Tennessee na década de 1860, na época de sua morte, em 1905, John William “Bud” Rogan tinha 2,67 metros de altura. Essa altura extraordinária o colocava bem longe na cauda direita da distribuição, tornando sua existência altamente improvável. Quão improvável? Isso pode ser estimado usando a fórmula para a distribuição normal. Com sua elegância típica, ela só precisa de um número para nos dar a resposta: a quantidade de sigmas entre a altura de Rogan e a altura média da população. Registros históricos⁵ mostram que, para homens da sua época e contexto, a altura média era de 1,70 metro, com um desvio-padrão de cerca de 7 centímetros. Então, ele se agigantava 97 centímetros acima do homem médio do seu tempo, o que representa mais de 13 desvios-padrão – ou “sigmas”. Inserindo esse número na fórmula relevante, descobre-se que não somente ele era 1 homem em 1 milhão, ou mesmo em 1 bilhão. Era 1 homem em 10^{44} , ou 100 milhões de trilhões de trilhões de trilhões, que excede por um gigantesco fator de aproximadamente 100 bilhões a quantidade de pessoas que já viveram até hoje.

Mais uma vez, nunca se deve esquecer que o extremamente inusitado pode acontecer, e acontece. Mas, como ocorreu com os icebergs de Viniar, não esperamos ver tais casos repetidamente. Na verdade, há pelo menos dezessete casos conhecidos de pessoas com alturas similares à de Rogan, entre elas Robert Wadlow (1918-1940), que era 5 centímetros mais alto e continua sendo a pessoa mais alta já registrada na história. A lição é clara: acreditar que tudo é normal é elaborar premissas que não se sustentam – com consequências que podem nos

deixar estupefatos ao lidar com extremos. Nunca devemos perder de vista o fato de que o teorema do limite central de Laplace vem com um pacote de termos e condições que, embora surpreendentemente flexíveis, não podem ser ignorados. Antes de nos voltarmos para a curva do sino em busca de informações, devemos sempre parar e perguntar se os dados são resultado plausível do efeito cumulativo de muitas variáveis atuando de forma mais ou menos independente. A confiabilidade do teorema de Laplace pode ser solapada pela falta de dados – e não há jeito fácil de saber se temos dados suficientes. Forçado a trabalhar nos confins do atrapalhado mundo real, o teorema se rebela advertindo sobre possibilidades ridículas que nunca veremos – ao mesmo tempo que falha em nos avisar acerca de extremos nos quais podemos tropeçar amanhã.

A declaração de Viniar anunciou o começo da crise e da recessão financeira global, cujo impacto será sentido ainda por muitos anos. E também provocou enorme debate sobre a premissa da normalidade. E não sem tempo: a evidência de que os mercados financeiros não seguem as restrições da curva normal tem sido clara há décadas para aqueles que têm olhos para ver.⁶ Em 2000, o altamente respeitado matemático das finanças britânico Paul Wilmott tentou avisar seus colegas *quants* acerca do perigo do que eles faziam: “Está claro que urge desesperadamente repensar tudo, se o mundo quiser evitar o derretimento do mercado causado pelos matemáticos. ... As premissas subjacentes aos modelos, tais como a importância da distribuição normal, a eliminação do risco, correlações mensuráveis etc., estão incorretas.”⁷

Isso não fez um pingão de diferença; as instituições financeiras fizeram apostas de risco cada vez maiores, ao mesmo tempo que mantinham os modelos matemáticos para encobrir sua exposição. “Enquanto a música estiver tocando, você tem de se levantar e dançar. Ainda estamos na dança”, declarou Chuck Prince, diretor executivo do Citigroup, no começo de 2007. Sua dança não duraria muito tempo. Na época, seu banco era o maior do mundo, com mais de US\$ 40 bilhões de exposição econômica na forma de CDOs, as Collateralised Debt Obligations, ou Obrigações de Débito Colateralizadas, um tipo de obrigação garantida por ativos como as hipotecas. A atração dos papéis como as

CDOs é a taxa de juros que pagam – muito mais alta que uma enfadonha letra do governo. Inevitavelmente, as melhores taxas de juros estão vinculadas a obrigações que apresentam o risco mais elevado de nunca serem honradas, garantidas pelos ativos menos confiáveis. O desafio consistia em decidir se a taxa de juros valia o risco.

Felizmente, as agências de avaliação de crédito estavam dispostas (em troca de remuneração) a usar sofisticados modelos matemáticos para quantificar a não confiabilidade – ou “risco de calote”, no jargão. Mas os modelos não eram nada sofisticados. Todos eles tinham embutidas curvas do sino – e, pior ainda, eram usadas para estimar o risco de eventos extremos. Ironicamente, para o ramo de negócios notório por fazer vigorar “termos e condições” para seus clientes, as instituições pareciam nem conhecer nem se preocupar com os “termos e condições” da curva do sino. No entanto, não é necessário um doutorado para desconfiar que muito provavelmente estavam prestes a sofrer uma séria quebra nos modelos de risco das CDOs.

Em poucos meses após a alegre declaração de Prince, os modelos de risco haviam revelado sua inadequação, e as CDOs começaram a ser descumpridas, ou a “sofrer calotes”, com índices catastróficos. O Citigroup se viu diante da falência, e precisou ser salvo por um resgate de US\$ 45 bilhões por parte do governo americano. E não foi só ele; no começo de 2008, a crise financeira global havia mostrado que os alertas de Wilmott não haviam sido (se é que foram alguma coisa) sombrios demais. “Eu venho cometendo um grande erro”, ele escreveu na época; “tenho sido sutil demais, ... é preciso gritar e bradar.” E detalhou suas advertências de forma mais clara e direta, afirmando que a falta de independência oculta nos modelos podia levá-los a “estourar drasticamente”. Aconselhou que se parasse imediatamente de usá-los. Isso não aconteceu. Papéis como CDOs e derivativos como CDSs (Credit Default Swaps, que são papéis negociados no mercado de renda fixa) são úteis – e lucrativos – demais para serem ignorados pelos financistas. Mas, em compensação, se forem empregados, devem se combinar com algo mais sofisticado que curvas do sino e pensamento

cobiçoso. Os reguladores estão clamando por modelos melhores, mas até hoje não parece ter havido grande melhora.⁸

Os reais condutores de mudanças são aqueles que estão sentados nas salas da diretoria dos gigantes financeiros. Se quisermos evitar a repetição da recente catástrofe, eles precisam ter mais conhecimento sobre o que seus homens de números tramam – e, se der errado, que sejam obrigados a encarar a música, em vez de tirar o corpo fora. Há sinais de que o recado talvez tenha começado a ser entendido. Quando os mercados do Tesouro dos Estados Unidos passaram por uma situação de 7,5 sigma em um único dia de outubro de 2014, o diretor executivo do JPMorgan, Jamie Dimon, disse aos acionistas que esses eventos deviam acontecer apenas a cada tantos bilhões de anos. Mas aí acrescentou um comentário significativo: como o mercado do Tesouro só existia há cerca de duzentos anos, “esse ‘deviam’ deveria fazê-los questionar a estatística, para começo de conversa”.⁹

Talvez tenha sido necessária uma calamidade para isso, mas parece que finalmente estamos nos movendo para além da curva do sino.

Conclusão

A curva do sino vem com “termos e condições” que frequentemente não podem ser atendidos no mundo real. Às vezes isso não tem muita importância. Mas se você está usando uma curva do sino para prever extremos, tome cuidado: abuse dos “termos e condições”, e você poderá provocar uma tempestade.

29. Irmãs feias e gêmeas malvadas

LAKE WOBEGON, Minnesota, é um lugar muito especial. Seu filho mais famoso é o contador de histórias bestseller Garrison Keillor, que vem cativando audiências com monólogos sobre sua cidade natal desde os anos 1970. E, como ele se delicia em explicar no encerramento de cada um de seus relatos, a cidade é um lugar onde “todas as mulheres são fortes, todos os homens são bonitões e todas as crianças são acima da média”. Muita gente vê isso como uma típica e caprichosa expressão do seu orgulho pelo lugar. Outros veem como uma grande pista da verdade sobre Lake Wobegon: o lugar não existe – porque essas crianças são uma impossibilidade.

Bem, nem tanto: é perfeitamente possível que todas as crianças de Lake Wobegon sejam, digamos, acima da média em relação a algum traço universalmente definido, como altura ou QI; todo corredor de 100 metros rasos na final olímpica é acima da média em corrida. Mas Keillor sugere que todas as crianças são acima da média *segundo qualquer parâmetro*, e isso é forçar a barra. Na verdade, se uma característica específica segue a curva do sino, há somente 50% de chance de que alguém escolhido ao acaso esteja acima da média. O inverso também é verdade, e isso tem uma implicação muito assustadora em termos de QI (que, como se constata, não segue uma curva do sino razoável): metade das pessoas no país tem inteligência abaixo da média. Tudo isso reflete uma peculiaridade da curva do sino: seu pico mostra não só onde se encontra o valor médio (a “média”), mas também o valor da *mediana*.

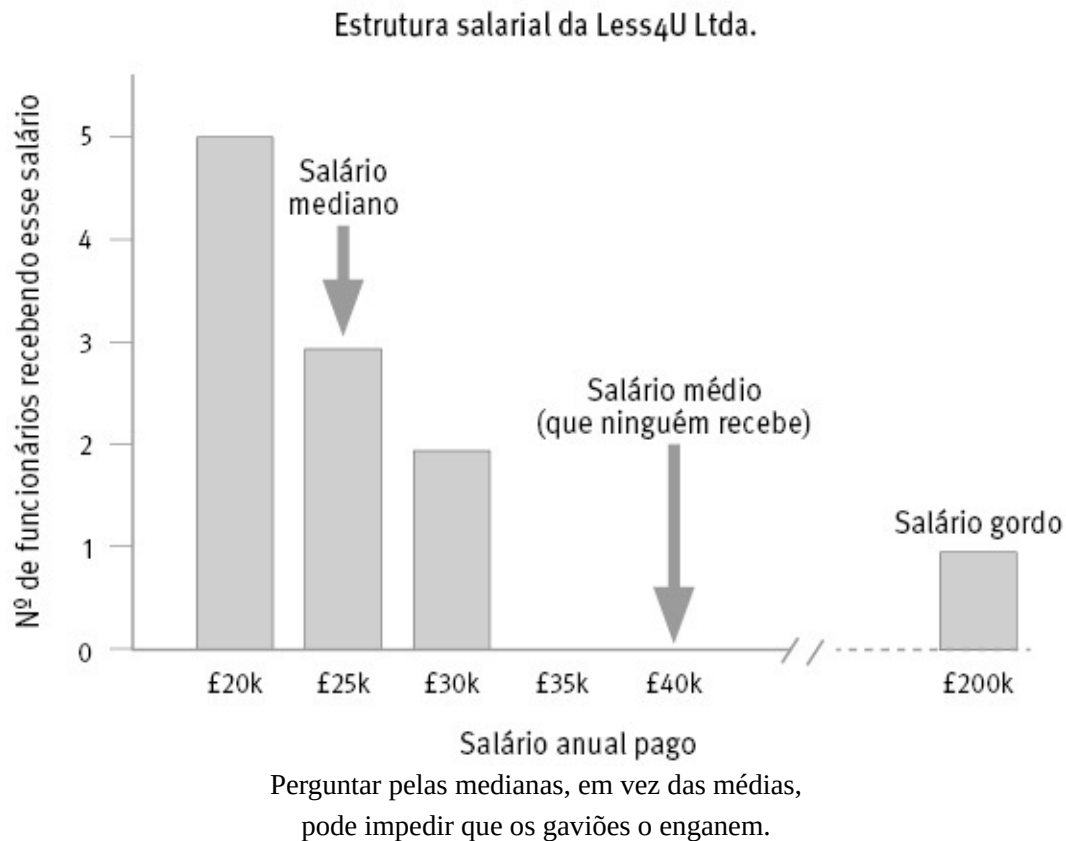
Como diz a palavra, a mediana é uma estatística resumida, um só número que sintetiza uma coleção de dados. Muitas vezes é vista apenas como um termo rebuscado para designar a mesma coisa que a média, mas ela é bem distinta e com muita frequência mais informativa. Para a familiaridade que ela tem, a média representa algo bastante esotérico: é a melhor estimativa do que se obterá

ao se retirar aleatoriamente um valor do meio dos dados. Isso é conveniente para características como altura ou QI, que seguem uma distribuição bem-arrumada, bastante *simétrica*, como a curva do sino, com quantidades iguais acima e abaixo do valor mais comum. Mas pode também se revelar barbaramente enganosa quando usada com fenômenos da vida real que não seguem distribuições tão bem-arranjadas.

Em contraste, a mediana é uma medida bastante robusta. Ela é definida como o valor que divide os dados ao meio, de modo que 50% de todas as medidas fiquem abaixo dela e 50% fiquem acima. Para dados que seguem a curva do sino, a mediana acaba sendo igual ao valor médio,¹ mas o que é importante em relação a ela é que desempenha o seu papel mesmo que os dados *não sigam* a curva do sino. De fato, eles podem seguir todos os tipos de distribuição menos bonitinha, e ainda assim dar uma mediana bem-definida que divide os dados igualmente em “altos” e “baixos”.

Isso torna a mediana especialmente útil se você desconfia de que alguma característica não segue realmente a curva do sino. Por exemplo, imagine que você esteja se candidatando ao emprego numa pequena empresa de cerca de uma dezena de pessoas que anuncia sua média de salários em torno de £40 mil. Isso parece impressionante... Até você perceber que os salários não estão distribuídos segundo a curva do sino ou, na verdade, segundo qualquer distribuição simétrica. Tampouco é provável que você obtenha nessa faixa um salário tirado ao acaso, que seria a média. Se ela for como a maioria das empresas, os salários são altamente enviesados, com a maioria recebendo quantias modestas, enquanto um punhado de gaviões ganha uma fortuna. A não ser que você esteja se candidatando a gavião, deveria pedir para saber o salário *mediano*. A diferença pode ser impressionante: no caso da Less4U (ver a seguir), *ninguém* ganha a cifra média, porque ela perde totalmente o sentido diante da enorme disparidade entre o salarião e as decepçionantes £25 mil da mediana. Em geral, sempre que a mediana vem radicalmente *abaixo* da média, como nesse caso, isso indica que a distribuição é fortemente enviesada para valores *mais*

baixos – sendo a média enganosamente inflada por pontos extremos, aqui, um salário gordo.



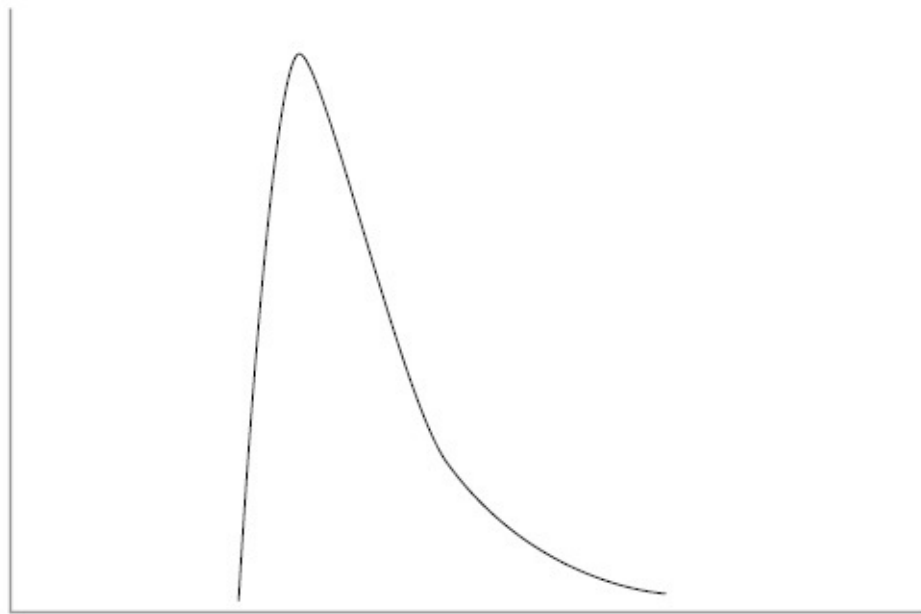
Distribuições enviesadas podem não ter um aspecto tão bonito quanto a curva do sino, mas elas são bem frequentes. De fato, os homens oferecem um exemplo excelente... no formato de seus pênis. Ou, para ser mais preciso, no tamanho: segundo uma pesquisa,² o comprimento médio do pênis é de 13,24 centímetros, mas o valor mediano é de 13,00 centímetros. Isso revela dois fatos intrigantes. Primeiro, mostra que a distribuição global de tamanhos de pênis é enviesada, no sentido de valores menores; segundo, que a maioria dos homens realmente tem pênis de tamanho *abaixo* da média. Outro exemplo de distribuição enviesada diz respeito à habilidade de dirigir. Muitos se declaram motoristas acima da média,³ crença com frequência desprezada como algo ridículo; realmente, ela tem sido atribuída a um efeito psicológico conhecido como *superioridade ilusória*. Contudo, mais uma vez, devemos aqui ter cautela para

não cair na armadilha de presumir a vigência de uma distribuição normal. No Reino Unido, pelo menos, os motoristas jovens têm probabilidade bem maior de se envolver em acidentes graves, apesar de comporem apenas pequena fração do número total de motoristas.⁴ Isso quer dizer que a distribuição da habilidade de guiar está distorcida no sentido de implicar que a maioria dos motoristas é melhor que a média – embora não esteja claro se a proporção de fato é tão alta quanto nós motoristas acreditamos. Em geral, porém, precisamos tomar cuidado ao desconsiderar afirmações aparentemente “estúpidas”, como “A maioria de X é melhor/pior que a média”. Distribuições enviesadas podem surgir por toda parte.

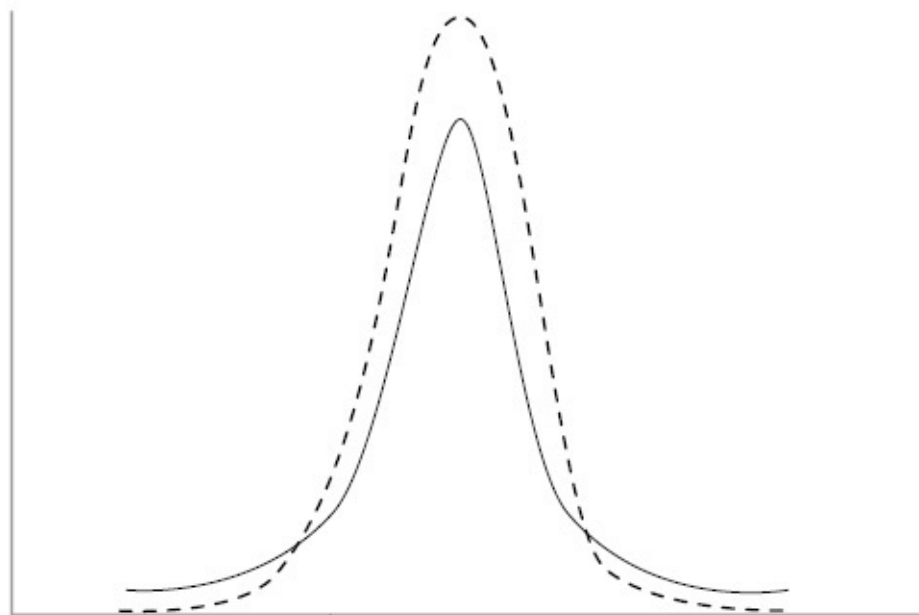
Decerto elas não são nada raras no mundo natural, aparecendo em todo lugar, desde meteorologia e ecologia até geologia. Isso acontece em parte porque os fenômenos da vida real obrigatoriamente se situam dentro de intervalos finitos. Tomemos as alturas: segundo a curva do sino, é possível ter pessoas com altura zero e até negativas, mas o senso comum sugere outra coisa. Os pesquisadores, portanto, muitas vezes são compelidos a torcer seus dados brutos (“transformá-los logaritmicamente”, é a expressão educada) para reduzir o punhado maior de uma das extremidades e forçá-lo dentro de algo mais com o formato de sino. Isso não é a trapaça que parece, e redundante alegar que os fenômenos se devem a influências aleatórias independentes que se *multiplicam*, em vez de simplesmente se somar.⁵ E os fenômenos “multiplicativos” são comuns em todas as ciências que estudam a vida, a química e a física. Na verdade, isso pode constituir excelente argumento para despir a curva do sino de seu enganoso apelido de “distribuição normal”, conferindo-lhe, em vez disso, o título de sua relação logarítmica menos renomada.⁶ Com sua falta de simetria, essa “irmã feia” da linda curva do sino carece de apelo estético, mas pode refletir melhor o mundo feio em que vivemos (ver Gráfico a seguir).

Ainda assim, aqueles que usam a curva do sino com negligência correm o risco de presenciar algo muito mais terrível que a mera perda de simetria. Podem se confrontar com as consequências verdadeiramente monstruosas do fracasso do teorema do limite central de Laplace. Apropriadamente, o mais antigo

vislumbre dessas consequências veio por intermédio de uma curva estudada pela primeira vez por matemáticos do século XVIII e conhecida como bruxa de Agnesi. Não está muito claro por que a curva veio a receber esse nome, mas ele parece adequado, considerando-se os efeitos demoníacos preditos sempre que ela está à espreita no meio dos dados. À primeira vista, parece exatamente uma curva do sino: um pico central com encostas graciosas descendo simetricamente de cada lado. Mas há algo que não é bem igual – e que fica claro quando se projeta uma curva do sino sobre ela (ver Gráficos a seguir).



A curva log-normal: mais feia que a curva do sino, embora talvez mais útil.



A bruxa de Agnesi (linha cheia) tenta enganar
você fazendo-se passar pela normal.

O pico da bruxa de Agnesi é mais acentuado, mais pontudo, porém, suas encostas são mais graciosas e relutam mais em sumir para os lados.⁷ Os matemáticos chamam essas curvas de “leptocúrticas”, da palavra grega para “leve arqueamento”, mas aquelas com as quais deparamos na vida real têm um nome bem menos lisonjeiro: “curvas de caudas grossas”. Isso é sintomático do fato de que, além das aparências, não existe nada de muito bonito na bruxa de Agnesi. Os dados que se conformam ao seu formato seguem aquela que agora é conhecida como distribuição de Cauchy, em honra a um fértil matemático francês do século XIX. E, apesar das semelhanças com a curva do sino, e seguindo uma fórmula muito mais simples, a distribuição de Cauchy é um ninho de víboras matemáticas. Primeiro, os dados que se conformam a ela se recusam a possuir um valor médio. Certo, é possível pegar, digamos, mil pontos de dados e calcular sua média somando-os e dividindo o resultado por mil, mas o resultado *não terá nenhum sentido*. O dado seguinte poderia ser tão diferente de todo o resto que mudaria totalmente a média. Você pode se ocupar com os dados na casa das dezenas e centenas quando, de repente – bumba! –, aparece o valor 51 319. Ao contrário dos dados que seguem a curva do sino, em que adicionar mais dados oferece uma estimativa melhor do valor médio, adicionar mais dados à

distribuição de Cauchy não faz diferença: tudo que se obtém são dados sempre mutáveis.

O mesmo acontece com qualquer tentativa de estimar o nível de variabilidade, conforme captado pelo desvio-padrão. Para a curva do sino, o desvio-padrão é refletido pelo grau em que a curva se espalha para cada lado do pico central. A curva de Cauchy claramente também se espalha, assim, o desvio-padrão não é zero. Mas tente estimar seu valor usando cem, mil ou 1 trilhão de pontos de dados, e você irá deparar com o mesmo problema que ocorre com a média: os resultados simplesmente se espalham por toda parte. Em outras palavras, a média e o desvio-padrão de Cauchy não são grandes, pequenos ou algo intermediário. Apesar do que sugere o formato da curva, eles simplesmente *não existem*.

Livros-texto de estatística e probabilidade dedicam pouco espaço à distribuição de Cauchy. Quando chega a ser mencionada, em geral é retratada apenas como maluquice matemática parecida com a distribuição normal, mas sem ser ela.⁸ No entanto, essa é exatamente a razão por que se deve conhecer melhor a distribuição de Cauchy: ela representa o cartaz anunciando o perigo de se assumir que tudo é normal. Em nenhum lugar é mais visível do que quando se tenta estimar as chances de obter resultados malucos. Estes são, por definição, inusitados, e, portanto, se situam nas caudas de muito baixa probabilidade, longe dos picos centrais da curva do sino ou da distribuição de Cauchy. Contudo, uma olhada rápida na superposição de uma sobre a outra mostra que elas não darão as mesmas respostas. As “caudas” mais grossas da distribuição de Cauchy sugerem que ela atribuirá chance mais alta que a da curva do sino aos resultados malucos. Contudo, quanto mais alta será ela? Isso requer cálculos cujos resultados são dados na Tabela a seguir.

PROBABILIDADE DA CURVA DO SINO 1 CHANCE EM...	PROBABILIDADE EQUIVALENTE NA DISTRIBUIÇÃO DE CAUCHY 1 CHANCE EM...	A CURVA DO SINO SUBESTIMA AS CHANCES POR UM FATOR DE...
20	7	3

100	9	11
1 000	11	91
1 milhão	16	62 500
1 bilhão	19	53 milhões
1 trilhão	23	43 bilhões

Se você confia na curva do sino para avaliar eventos raros, prepare-se para ter um choque.

As diferenças entre as previsões são realmente chocantes, especialmente considerando as aparentes semelhanças das duas curvas. E isso mostra o perigo de se assumir sem mais nem menos que os dados que se encaixam em algo *parecido* com uma curva do sino realmente são normais. Isso vale em particular para os eventos raros. Por exemplo, um evento esperado em média 1 vez em cada 1 bilhão de anos numa distribuição “normal” poderia aparecer 1 vez em 19 anos se seguir a distribuição de Cauchy. De repente, não é surpresa que sujeitos do tipo de Jamie Dimon, do JPMorgan, tenham presenciado um movimento de mercado de 1 vez em 1 bilhão. Até o relato de David Viniar, de que vivenciou em poucos dias eventos que nunca deveriam ter acontecido na história do Universo, não parece tão extraordinário.⁹ Ou, pelo menos, não se deveria acreditar que a distribuição de Cauchy se aplicasse a esses eventos. Mas será que isso é mesmo plausível? Podem eventos da vida real seguir algo tão excêntrico como a distribuição de Cauchy, com sua bizarra relutância em fornecer até valores médios?

Dadas as fortunas que por aí viajam, não é surpresa que os pesquisadores venham tentando, durante décadas, encaixar distribuições nos dados financeiros. E considerando a tendência de ver a curva do sino em todo lugar, os primeiros estudos alegavam que os movimentos dos preços das ações de fato seguiam seus ditames. Todavia, já em meados dos anos 1960, estava claro que isso não passava de um desejo esperançoso. Numa celebrada tese de doutorado publicada enquanto ainda estava na casa dos vinte anos, o economista americano Eugene Fama, depois ganhador do Prêmio Nobel, mostrou que há mudanças extremas

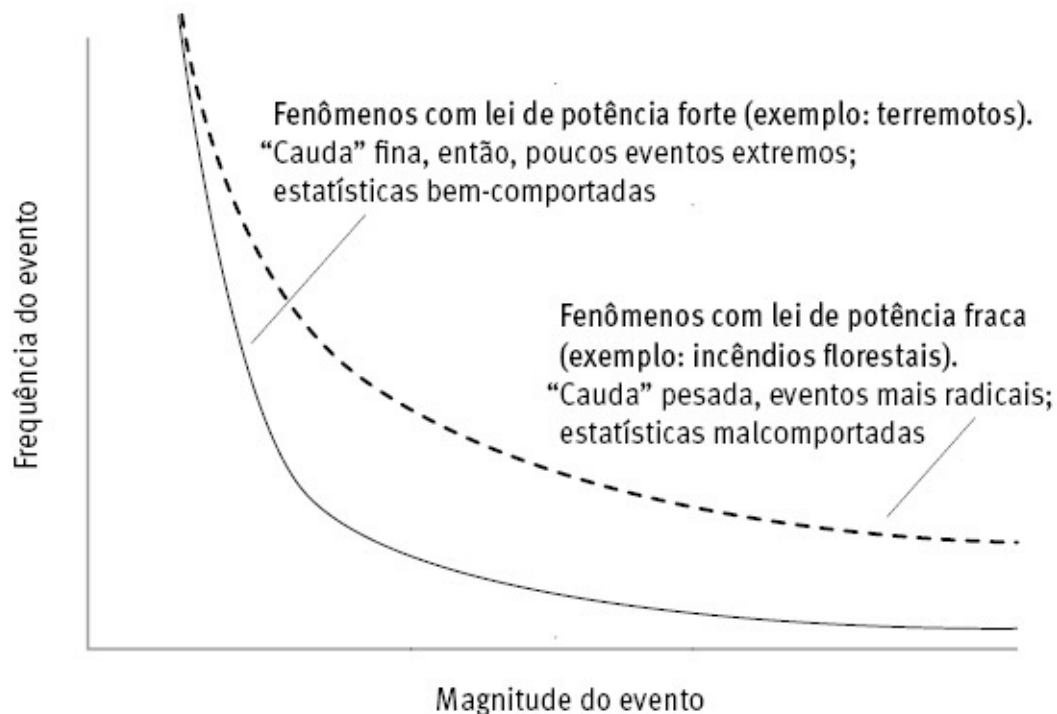
demais nos preços das ações. E isso dava à distribuição um pico central mais pontudo e caudas mais grossas do que se espera numa curva do sino¹⁰ – em outras palavras, aquilo estava mais para uma distribuição de Cauchy. No entanto, Fama descobriu que a coisa era mais interessante ainda. O melhor encaixe vinha de se empregarem curvas pertencentes a toda uma família de distribuições da qual Cauchy e a curva do sino são apenas casos especiais. Conhecidas enigmaticamente como distribuições de Lévy-estáveis,¹¹ elas podem ser benignas como a curva do sino ou muito doidas, como a de Cauchy.¹² Fama descobriu que os movimentos dos preços das ações têm uma distribuição que se encontra em algum ponto intermediário.

O que não estava claro era por quê. Obviamente, seu comportamento devia violar pelo menos um dos “termos e condições” do teorema do limite central subjacente à curva do sino – e a candidata óbvia era a independência. Afinal, todo mundo sabe que os investidores são como carneiros, todos compram “dicas quentes” ou vendem “furadas”. No entanto, Fama descobriu que o preço das ações num determinado dia era mais ou menos independente de seu valor até dezesseis dias antes. Então, se a premissa de independência estava certa, o que mais podia dar errado? Fama achou a pista na pura violência dos movimentos do mercado de ações. Como a distribuição de Cauchy, eles têm um desvio-padrão patologicamente grande, que pode ser ao mesmo tempo súbito e imenso. Esse comportamento não pode ser abarcado pelo teorema do limite central, e sua curva do sino fica distorcida em algo mais pontiagudo, de caudas mais grossas – e absolutamente mais perigoso.

Isso não deveria ser surpresa para quem já experimentou a montanha-russa financeira das últimas décadas. O que deveria nos scandalizar a todos é que tudo era sabido mais de meio século antes. Estudiosos como Fama mostraram que, enquanto os preços de um dia particular podem seguir os ditames da curva do sino, ainda assim eles são capazes de se movimentar de forma terrivelmente abrupta. Como tal, confiar na curva do sino para estimar o risco de determinada perda é em si arriscadíssimo, quase criminalmente irresponsável. A despeito de

tudo isso, as curvas do sino continuaram embutidas nas estimativas de risco, até mesmo no setor financeiro.

A distribuição de Cauchy é a gêmea malvada da curva do sino, capaz de se fazer passar pela irmã mais benigna, com seu pico elegante e caudas graciosas, mas também de se comportar muito mal. Contudo, ela não está sozinha. Assim como suas parentes próximas na família Lévy-estável, os traços mais desagradáveis de Cauchy são compartilhados pelas chamadas distribuições de lei de potência – descobertas ali à espreita, numa legião de fenômenos da vida real, desde terremotos e incêndios florestais até riqueza pessoal. Matematicamente, leis de potência são muito mais simples que a curva do sino, mas fazem o mesmo serviço de ligar o tamanho de um fenômeno à sua prevalência (ver Gráfico a seguir). Suas origens parecem ser tão variadas quanto os fenômenos que descrevem;¹³ no entanto, todas compartilham a mesma aparência básica: inexistência de um pico central, parecem mais a beirada de um penhasco que se precipita e depois se alonga numa comprida cauda refletindo a característica básica dos fenômenos que descrevem: maior significa mais raro. Peguemos os terremotos: enquanto a maioria deles são fracos demais até para serem notados, alguns são desastrosos – e poucos são devastadores. Registros históricos têm permitido aos sismólogos identificar isso com maior precisão, levando-os à chamada relação de Gutenberg-Richter, mostrando que há dez vezes menos abalos com magnitudes Richter entre 6 e 7 do que entre 5 e 6, e ainda dez vezes menos entre 7 e 8. Um declínio tão drástico é típico da lei de potência, e nesse caso ela é bastante simples e bem-comportada (ver Gráfico a seguir).



Estranhamente, quanto mais fraca a lei de potência, maior o ferrão na cauda.

Isso ao menos permite que se extraiam dos dados valores estáveis para os tamanhos médios de terremotos – o que é mais do que se pode esperar de uma curva de Cauchy. Mas nem todos os fenômenos que obedecem à lei de potência são tão benignos: erupções solares, incêndios florestais e conflitos humanos têm mostrado que todos eles seguem distribuições de lei de potência para as quais nem tamanhos médios nem mesmo intervalos plausíveis podem ser estimados de maneira confiável. Essas leis de potência são simplesmente tão relutantes quanto a curva de Cauchy, e nos pregam suas peças estatísticas. Isso tem sérias consequências práticas. Por exemplo, como se pode ter certeza de enfrentar o risco de grandes incêndios florestais quando mesmo o seu tamanho médio é tão difícil de se estimar? A existência de curvas de potência também ameaça a confiabilidade de informações sobre fenômenos erroneamente considerados seguidores da curva do sino.¹⁴ Os pesquisadores correm o risco de calcular as estatísticas básicas como médias, inconscientes de que as leis de potência que dirigem os dados podem tornar esses cálculos sem sentido. Como veremos, elas também podem mandar para o lixo métodos de analisar dados e encontrar

padrões, bem como minar tentativas de replicar descobertas que nelas se baseiam.

Em suma, essas distribuições patológicas têm o poder de solapar os métodos da própria ciência. Como tal, elas nos apresentam um desafio fundamental: aceitamos sua existência e aprendemos a conviver e a trabalhar com elas, ou devemos continuar a confiar em modelos de realidade que são simples, elegantes e errados?

Conclusão

O mundo real abriga uma legião de fenômenos que parecem inteiramente normais, mas são qualquer coisa, menos normais. Pior ainda, dados sobre esses monstros matemáticos podem fazê-los parecer inteiramente benignos. Ainda assim, a menos que sejam identificados e abordados cuidadosamente, eles podem fazer troça das nossas tentativas de compreendê-los.

30. Até o extremo

A IDEIA DE JOGAR a culpa pela crise global em uma pessoa só raramente faz sentido. Mas no caso do colapso financeiro de 2007-08 um nome veio à tona mais que qualquer outro: Alan Greenspan. De 1987 até alguns meses antes de a crise eclodir, Greenspan foi presidente do US Federal Reserve – o Banco Central dos Estados Unidos, o FED; em outras palavras, ele era chefe do sistema bancário central da maior economia do mundo. Durante essa época, dizem seus críticos, instituiu um regime de regulação financeira cada vez mais frouxa, guiado por uma crença quase religiosa nos benefícios do livre mercado. O resultado foi uma ganância sem freios, níveis insanos de alavancagem e riscos e um desastre de muitos trilhões de dólares.

Não há escassez de evidências contra Greenspan – nem mesmo depois de seu mea-culpa perante uma comissão do Congresso, em 2008, no qual admitiu estar num “estado de chocada descrença” em relação ao que acontecera sob sua vigilância. Contudo, ele merece crédito por estar entre os primeiros a manifestar preocupação sobre as perigosas premissas à espreita nos modelos de risco usados nas finanças. Falando numa conferência para presidentes dos bancos centrais em 1995, Greenspan advertiu sobre o “uso inapropriado” da curva do sino, com sua tendência a subestimar as chances de eventos extraordinários. A audiência começava a ficar desconfortavelmente familiarizada com esses eventos. Em fevereiro daquele ano, o mais famoso banco mercantil do mundo, o Barings, tinha desabado depois de perder mais de £800 milhões (o equivalente a cerca de £1,5 bilhão de hoje) como consequência das operações de um único negociante chamado Nick Leeson. Então o Daiwa Bank do Japão descobriu um buraco igualmente vasto em suas contas, fruto das atividades de outro especulador inescrupuloso. Para Greenspan, a lição era clara: os bancos centrais tinham de ver a si mesmos como companhias seguradoras capazes de prover cobertura até

no caso de catástrofes. E isso, insinuava Greenspan, significava lançar mão com maior frequência de um kit de ferramentas matemáticas que cada vez mais servia de base para a indústria de seguros, com efeito impressionante: a teoria dos valores extremos (TVE).

A ideia de que os extremos eram mais do que pontos radicais em distribuições familiares fora reconhecida no começo do século XVIII por um dos pioneiros da teoria da probabilidade, Nicolau Bernoulli. Entretanto, apesar de sua óbvia importância, foram necessários mais duzentos anos para que a teoria por trás deles emergisse. Na década de 1920, o sempre brilhante Ronald Fisher, com seu ex-aluno Leonard Tippett, provou que eventos extremos seguem distribuições próprias e especiais.¹ Estas foram posteriormente combinadas numa única fórmula, conhecida como distribuição generalizada de valores extremos (GVE), cujo formato pode ser sintonizado usando dados sobre eventos radicais. As curvas resultantes são um tanto estranhas matematicamente, mas ainda assim refletem a ideia de senso comum de que, quanto mais extremo o evento, menos provável ele é. O mais importante, porém, é que as previsões detalhadas podiam ser totalmente diferentes daquelas que emergiam a partir da curva do sino.

Com os modelos de negócios constantemente sob ameaça de eventos extremos, as companhias de seguros passaram a estudar assiduamente a TVE. Durante anos, analistas aferiram o risco provável apresentado por várias formas de desastre usando regras práticas empíricas tais como a regra “20-80”, que afirma que 20% dos eventos graves contribuem para mais de 80% do total de indenizações.² Em meados dos anos 1990, o matemático das finanças Paul Embrechts e seus colegas no Instituto Federal Suíço de Tecnologia (ETH), em Zurique, resolveram checar a validade dessas regras com a TVE. Descobriram que a regra “20-80” funciona bem para muitos setores de seguros, mas, quando falha, falha muito feio. Usando a TVE para estudar dados passados acerca de pedidos de ressarcimento, o grupo descobriu que uma regra “0,1-95” aplica-se a danos causados por furacões. Em outras palavras, enquanto todos os furacões são um desafio em potencial, a verdadeira ameaça vem apenas de 1 em 1 000, que pode devorar 95% de toda a cobertura da companhia de um só golpe. Tais

descobertas permitiram às seguradoras otimizar sua cobertura de risco, ampliando a gama de ameaças que podem cobrir com prêmios sensatos, beneficiando tanto a si próprias quanto a seus clientes.

A TVE agora é usada para proteger aqueles cuja vida fica sob risco durante essas calamidades naturais. Um país efetivamente apostou seu futuro nas previsões da teoria. Em fevereiro de 1953, uma tempestade gigantesca assolou a costa do mar do Norte na Europa. As inundações resultantes mataram mais de 2 500 pessoas, inclusive 1 800 na Holanda, cujas seculares defesas contra o mar foram sobrepujadas. Determinado a impedir uma repetição para as gerações vindouras, o governo holandês convocou uma comissão de especialistas para projetar defesas marítimas capazes de atender ao padrão sem levar o país à falência. A comissão estimou que defesas costeiras com cerca de cinco metros acima do nível do mar bastariam. Mas era possível confiar nesse número? Registros mostravam que o evento de 1953 não fora excepcional: enchentes severas tinham atingido a Holanda dezenas de vezes ao longo do último milênio. Em 1º de novembro de 1570, Dia de Todos os Santos, o país foi devastado por uma tromba-d'água com mais de quatro metros – mais de quinze centímetros acima do evento de 1953 –, resultando em dezenas de milhares de mortos. As preocupações levaram o governo holandês a encarregar uma equipe liderada por Laurens de Haan, perito em TVE, da Universidade Erasmus, em Rotterdam, de avaliar o padrão de cinco metros. Usando dados históricos de enchentes, a equipe estabeleceu a curva TVE que levava em conta as inundações extremas passadas – e a extrapolou para o futuro. Eles descobriram que as recomendações originais seriam aquelas mesmas por muitos séculos.

Se elas vão ou não se manter, isso fica para ser verificado depois; como descobrimos, nunca é sensato depositar fé cega em modelos matemáticos, não importa quão sofisticados eles possam parecer. Decerto há base para preocupações sobre a confiabilidade da TVE, porque – assim como na curva do sino – seus impressionantes poderes vêm com uma longa lista de “termos e condições”. Uma questão-chave diz respeito aos próprios dados usados para estabelecer a melhor curva TVE para aquela função. Como o nome sugere,

precisamos de exemplos de casos extremos – mas o que conta como “extremo”? Ao pescar registros históricos, é necessário estabelecer algum tipo de limiar, mas onde? Fixá-lo baixo demais deixará de fora muitos casos duvidosos, tornando a curva pouco acurada; um limiar alto demais tornará o conjunto de dados tão delgado que a curva se torna confusa e imprecisa. Depois há o problema que infecta a curva do sino: o que está guiando os dados observados – essas influências são independentes e imutáveis? Dada a evidência da mudança climática ao longo dos séculos, essas poderiam parecer premissas questionáveis para se adiantar acerca de terremotos, enchentes e furacões.

Tampouco a TVE está livre daquelas pragas mais bizarras e perniciosas: médias e intervalos instáveis. Como no caso das leis de potência e curvas semelhantes às de Cauchy, alguns tipos de distribuição TVE são profundamente tendenciosas. Pesquisa usando dados da vida real sobre perdas extremas sofridas por bancos descobriu que as curvas resultantes frequentemente não têm intervalos e valores médios bem-definidos.³ Isso torna as estimativas de risco instáveis. A adição de apenas mais alguns pontos de dados muda totalmente a cifra de risco e o tamanho do provável sinistro.⁴

Claro que não será fácil assumir a proposta de Greenspan e recorrer à TVE nos modelos financeiros. Hoje há esforços no sentido de resolver esses problemas – estimulados pelo fato de que, enquanto a TVE ainda é uma “obra em progresso”, ela é mais fácil de errar pelo lado da cautela que a curva do sino. A maior barreira para sua aceitação podem ser as próprias instituições financeiras. Depois do colapso, elas agora são obrigadas pelos reguladores a dispor de reservas capazes de cobrir situações em que negócios e empréstimos vão mal, para que nunca mais precisem de uma injeção de resgate. Calcular o tamanho dessas reservas é um desafio duro em termos de modelagem de risco. Mas está claro que, se o cálculo for feito usando TVE, as reservas se mostrarão substancialmente maiores que as exigidas quando se emprega a curva do sino.⁷ O problema é que os bancos não se mostram muito dispostos a manter quantias imensas paradas nos cofres para fazer face a um dia de chuva – e os reguladores lhes permitem escolher o método para fazer suas somas.

VIDAS EXTREMAS – E SEQUÊNCIAS DE DERROTAS

Desde a década de 1950 a expectativa de vida típica aumentou de cerca de 45 anos para mais de setenta, no mundo todo, e agora excede os oitenta em muitos países desenvolvidos. Essa tendência não pode continuar para sempre, claro, mas onde irá parar? Será que aquilo que hoje sabemos sobre longevidade dos homens pode ser usado para estimar o período máximo da vida humana? Na Universidade Erasmus, em Rotterdam, Laurens de Haan e seus colegas examinaram os registros de duração de vida para o “velho mais velho”, e então aplicaram TVE a fim de extrapolar para o período de vida humana definitivo. Eles acabaram com um número de mais ou menos 124 anos.⁵ Na época, a pessoa mais velha já registrada ainda estava viva: Jeanne Calment, de Arles, França, que se recordava de ter conhecido Vincent van Gogh aos treze anos. Ela morreu em 1997, com 122 anos – apenas dois a menos que o limite superior estabelecido usando TVE, o que hoje parece valer por mais alguns anos.

Ainda que a teoria por trás da TVE seja complexa, uma versão simples funciona para uma das situações extremas mais dolorosas da vida: longas sequências de derrotas. A fórmula resultante⁶ tem algumas implicações surpreendentes. Por exemplo, se lançarmos uma moeda 50 vezes, devemos nos preparar para ver sequências de cara (ou coroa) cerca de 5 vezes seguidas, com uma variação de mais ou menos 2. Elas são muito mais longas do que a maioria das pessoas espera – e ajuda a pôr as sequências perdedoras em perspectiva. Também lança alguma luz sobre uma famosa sequência perdedora vivida pelo palpiteiro de corridas de cavalos britânico Tom Segal na revista *Racing Post*. Usando o pseudônimo de Pricewise, Segal tem reputação de recomendar cavalos inimagináveis, com proporções relativamente altas nas apostas. Esses cavalos não têm probabilidade de ganhar com muita frequência, mas, quando ganham, ganham bonito. Em 2011, Segal teve uma sequência de 26 palpites errados consecutivos – levando muitos de seus seguidores a se preocupar com a possibilidade de ele ter perdido a mão. No entanto, a TVE mostra que, para os tipos de palpites improváveis que Segal dá, uma sequência de 32 apostas perdedoras seria inteiramente normal no decorrer de um ano. A sequência ruim terminou algumas semanas depois, e Segal prosseguiu produzindo impressionantes 20% de retorno de investimento para aqueles que conservaram a fé nele.

Decidirão eles não correr o risco de serem novamente pegos no meio da tempestade e trocar as atraentes e esbeltas caudas da curva do sino pelas caudas gordas e onerosas da TVE? Considerando que tivemos pelo menos cinco grandes crises financeiras desde a sugestão de Greenspan, em 1995, só se pode dizer uma coisa: não fique muito certo disso.

Conclusão

Num mundo assolado por extremos que vão desde um clima maluco até convulsões financeiras, a teoria dos valores extremos pode transformar registros históricos em informações sobre quanto as coisas podem ir mal. Assumir que futuro será igual ao passado é arriscado – mas se você acha que isso é perigoso, tente as adivinhações.

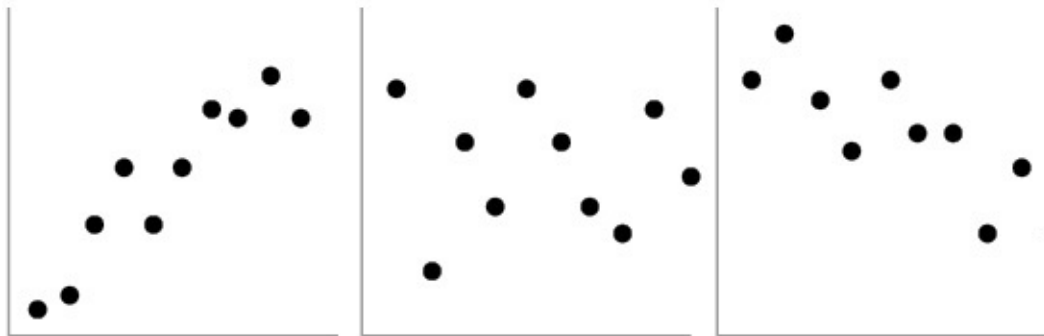
31. Assista a um filme de Nicolas Cage e morra

TODOS OS CIENTISTAS QUEREM fazer descobertas que mudem nossa visão da vida, do Universo e até acerca da natureza da realidade. A maioria precisa se contentar com algum insight diante do qual as pessoas param e notam. Por esse padrão, Tyler Vigen tem tido um êxito brilhante. Suas descobertas, relatadas no mundo inteiro, são empolgantes pelo caráter inesperado e pela quantidade. Até hoje, ele revelou dezenas de milhares de impressionantes insights, e ainda não parou. Ou, para ser mais exato, seu computador não parou.

Pois não é Vigen que está fazendo as descobertas; ele deixa isso para o seu computador, que programou para fazer exatamente o que os cientistas vêm fazendo há décadas: varrendo dados para descobrir como uma variável muda sob a interferência de outra. Essa é uma técnica que tem levado os cientistas a uma legião de descobertas, desde elos entre exposição a radiação e risco de câncer até a conexão entre as propriedades das estrelas e a expansão do cosmo. O computador de Vigen aplica os mesmos métodos, analisando os dados em busca de variáveis “altamente correlacionadas”. Isto é, busca conjuntos aleatórios de dados e aplica uma fórmula que cospe os chamados “coeficientes de correlação”. Estes podem variar de +1 – quando altos valores de uma variável correspondem a altos valores da outra –, passando por zero, quando não há padrão nenhum, até -1, quando altos valores de uma correspondem a baixos valores da outra e vice-versa (ver a seguir).¹

O computador de Vigen busca conjuntos de dados que, quando emparelhados entre si, produzam coeficientes de correlação próximos desses extremos. Isso porque é o que se espera encontrar se realmente houver uma ligação forte entre duas variáveis. Em contraste, coeficientes de correlação próximos de zero são sintomas de ausência de qualquer relação; portanto, não há nada de empolgante

acontecendo. Automatizando todo o processo, Vigen criou uma máquina de descobertas.

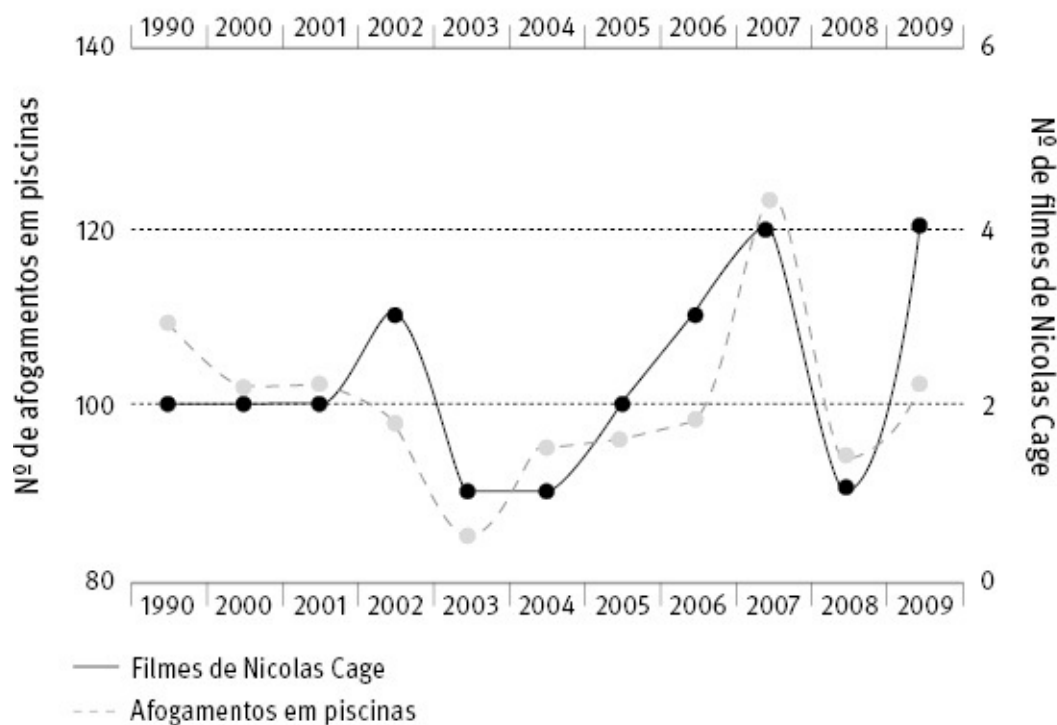


Três graus de correlação: $+0,85$; $0,0$; $-0,85$.
Todos podem ser importantes – ou disparates.

O que a máquina está descobrindo deve certamente mudar a nossa visão, mas não da realidade, e sim da confiabilidade de muitos anúncios de descobertas dignos de notícias e que se baseiam na mesma técnica. Vigen não é cientista; na época em que este livro é escrito, ele é aluno de graduação em direito em Harvard. Contudo, deixa seu computador caçador de correlações solto nos pastos ricos em dados da web e publica os resultados em seu site. Assim, ele está fornecendo um constante lembrete dos perigos de se aplicar irresponsavelmente um dos mais populares, porém mal utilizados, conceitos da ciência. Desde que foi posto para funcionar, o computador de Vigen descobriu uma legião de correlações absolutamente malucas. Tomadas superficialmente, elas sugerem que Nicolas Cage devia ser impedido de atuar em filmes, pois estes estão vinculados a mortes em piscina (coeficiente de correlação $+0,67$), e que os Estados Unidos deveriam banir a importação de carros japoneses, pois eles estão associados a suicídios por desastres de automóvel (coeficiente de correlação $+0,94$).

Entre as principais sacadas reveladas pelo computador de Vigen está uma que diz não ser boa ideia comer queijo como última refeição da noite, pois o consumo per capita do artigo está fortemente correlacionado à morte por sufocamento entre os lençóis ($+0,97$). Se você tem problemas nos seus relacionamentos, talvez queira também considerar se mudar para uma área que consuma relativamente pouca margarina: o computador de Vigen revelou que o

consumo per capita do produto está altamente correlacionado às taxas de divórcio – pelo menos, no estado americano do Maine.



Por que você deve evitar piscinas quando se lança um filme de Nicolas Cage.

Tudo muito divertido, e não é surpresa que o site de Vigen relatando essas “descobertas” tenha mais de 5 milhões de acessos. Afinal, elas parecem morbidamente as remissantes descobertas que encontramos com tanta frequência na mídia, acompanhadas de expressões como “Segundo os cientistas”. Seria bacana pensar que esses absurdos não teriam possibilidade de ganhar corpo entre os cientistas sérios, mas Vigen tem uma lição ainda mais importante para nós. A maioria das suas “descobertas” são estatisticamente significativas – o teste-padrão básico usado em pesquisa para avaliar se um achado é mais do que só uma casualidade sem sentido.² Sob esse aspecto, para manter essa coisa fora da bibliografia séria de pesquisa, as técnicas nas quais a maioria dos pesquisadores se baseia são frágeis demais. Temos de olhar para além delas.

A não plausibilidade pura e simples é a maneira mais óbvia de fazer isso. Nada além da não plausibilidade impede que a maioria das correlações seja levada a sério (por exemplo, importações americanas de petróleo da Noruega e motoristas mortos por trens – que tem o coeficiente de correlação extremamente significativo de +0,96). Outras correlações se estilhaçam e são consumidas pelo fogo no momento em que se examinam os números concretos por trás dos números brutos. Tomemos o exemplo da letalidade dos filmes de Nicolas Cage. Ele é um sujeito que trabalha duro e tem aparecido em vários filmes por ano, durante mais de uma década, mas até ele teve de se empenhar para fazer mais de três por ano. Em outras palavras, sua produção tem sido constante. Sob esse aspecto, ele é páreo para as ações da Sombria Ceifeira nas piscinas americanas. Durante a década de dados usados pelo computador de Vigen para achar uma correlação, houve cerca de cem fatalidades por ano, mas nunca menos que 85 ou mais que 123. No entanto, por acaso, esses picos ocorreram nos dois anos em que Nicolas também fez a quantidade mínima e máxima de filmes. Como o conjunto de dados é tão pequeno, a coincidência desses dois conjuntos de pontos extremos sobrepuja a frágil evidência nos outros valores mais ou menos constantes – e nós acabamos por achar que Nicolas Cage e a Sombria Ceifeira agem em conjunto (havendo, como é o caso, um coeficiente de correlação assustadoramente apropriado de +0,666). Esses “pontos atípicos” são conhecidos por criar e quebrar correlações quando há poucos dados disponíveis. Com frequência são tratados como uma prole de “erros experimentais” ou outra mancada, e simplesmente eliminados num processo eufemisticamente chamado “limpeza de dados”. No caso dos conjuntos de dados Cage/afogamentos, essa limpeza corresponde à metade do coeficiente de correlação, que também passa a ser não significativo. Contudo, na pesquisa científica de verdade, justificar essa eliminação nem sempre é simples. Pontos atípicos podem ser inteiramente genuínos quando se lida com fenômenos com comportamento de lei de potência, como fenômenos climáticos ou fatores econômicos.³

Nicolas Cage, claro, não tem nada a ver com nada, mas nem todas as “descobertas” de Vigen são ridicularizadas e dispensadas com tamanha

facilidade. Temos certeza, por exemplo, de que não há nada na correlação entre receita total gerada pelos campos de golfe nos Estados Unidos e a quantidade de dinheiro que os americanos gastam em esportes de espectadores (+0,95)? Talvez esse seja um reflexo do fato de que as pessoas que assistiram aos jogos de golfe tenham vontade de jogar. Ou talvez as pessoas que jogam golfe sejam chegadas a esportes em geral? A simples força da correlação não nos diz sequer se a relação é genuína; como diz o velho ditado, correlação não é causalidade. E tampouco a significância estatística diz alguma coisa sobre a “significância” real de uma correlação, a despeito do que muitos pesquisadores parecem pensar.

A significância estatística, lembremos, mede apenas as chances de se ter uma correlação pelo menos tão expressiva *presumindo-se* que seja mero acaso; não diz nada a respeito da veracidade ou não da premissa. Como vimos muitas vezes, responder a essa pergunta requer métodos bayesianos – e aqui eles trazem a vantagem extra de nos permitir considerar como fator nossas crenças a priori sobre a correlação. A princípio, isso pode ajudar a dar uma ideia sobre as chances de a correlação ser casual. Todavia, ainda é algo traiçoeiro, porque a correlação pode de fato ser real, crível, e mesmo assim ser alarme falso. Ela pode ser produto de uma “confusão” oculta – algum intermediário que ligue duas variáveis desconectadas entre si. Os casos de sérias queimaduras de sol estão, sem dúvida, significativamente correlacionados às vendas de óleo de bronzear – e também de sorvetes e refrescos gelados. Será que isso significa que estes últimos causam queimaduras de sol? Claro que não. Há um fator de confusão – um “confundimento” – não tão oculto conectando todos eles: o Sol.

Os resultados da confusão podem ser divertidos. Ninguém sabe muito bem quando ou como surgiu a ideia de que as cegonhas trazem os bebês, mas ela adquiriu status lendário entre os estatísticos, e vários estudos revelaram uma forte e estatisticamente significativa correlação entre populações de cegonhas e nascimentos em vários países. Uma explicação em potencial é o fator de confusão da área de terra – que está correlacionado tanto com populações de cegonhas quanto com taxas de natalidade.⁴ No entanto, os efeitos da confusão

nem sempre são tão interessantes. A menos que sejam identificados e corrigidos, podem acabar orientando as políticas públicas.

Fumar maconha tem se vinculado a inúmeros riscos para a saúde, e mesmo aqueles que nunca tocaram na droga sabem que ela deixa você meio abobado. A confirmação veio em 2012, num estudo publicado por uma respeitada revista que descobriu a ligação clara entre dependência da *cannabis* ao longo do tempo e perda de QI.⁵ Cientes da necessidade de evitar serem iludidos por elementos de confusão, os pesquisadores levaram em conta fatores como uso de álcool e drogas pesadas, mas o efeito permaneceu: aqueles que tinham adquirido o hábito na adolescência, tornando-se usuários contumazes e persistentes, perderam oito pontos de QI no fim da casa dos trinta anos. Mas espere aí – de qualquer modo, as pessoas não ficam mais esquecidas com o tempo? Isso é possível, e os pesquisadores também cobriram esse aspecto, comparando seu universo com pessoas de idade similar que jamais usaram *cannabis* (estranhamente, seus QIs na verdade aumentaram ligeiramente). Apesar de tudo, porém, os pesquisadores deram de cara com o problema habitual ao lidar com fatores de confusão: quanto mais esses fatores são desnudados, mais dados acabam excluídos da análise final. Tendo começado com mais de 1 000 pessoas no grupo de estudo original, os pesquisadores acabaram com apenas poucas dezenas livres da confundidora influência do álcool e abuso de drogas pesadas. E, como a equipe admitiu, esses dificilmente são os únicos fatores de confusão. Mesmo assim, ao confirmar “o que todo mundo sabe” sobre quem curte o barato de um baseado por longo tempo, o estudo recebeu uma enorme cobertura da mídia.

Entretanto, em poucas semanas, suas conclusões eram questionadas, por terem falhado em levar em conta outros fatores de confusão. Um deles é um intrigante fenômeno envolvendo escores de testes de QI crescentes observados em muitos países desde os anos 1930. Conhecido como efeito Flynn, o motivo de as pessoas que vivem hoje serem tão mais “inteligentes” que seus avós (ou, pelo menos, elas se saem melhor em testes de QI) ainda é debatido, mas uma possibilidade – respaldada pelo descobridor que dá nome ao efeito – é que vivemos cada vez mais em ambientes ricos em tarefas do tipo testes de QI, e

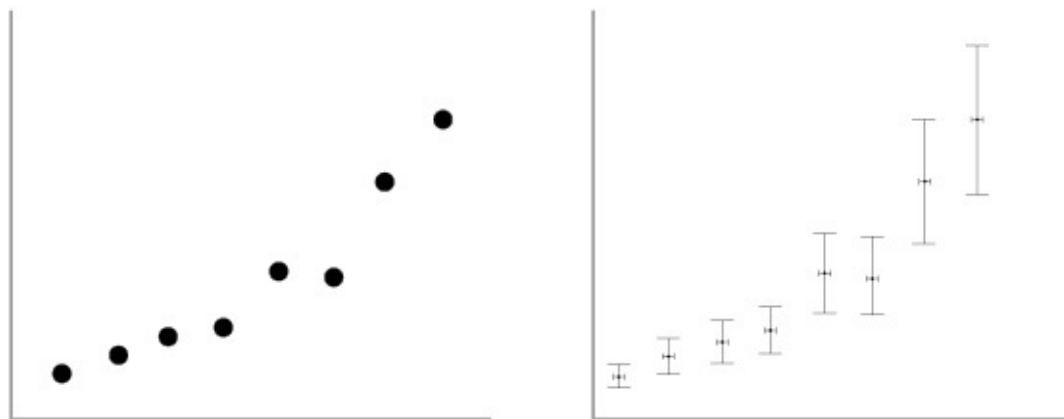
aqueles que são especialmente bons nesses testes descobrem-se em situações que lhes apresentam ainda novos desafios, o que amplia mais ainda o efeito. Qualquer que seja a explicação, o efeito Flynn claramente precisa ser levado em conta em qualquer estudo que focalize mudanças de QI com o tempo; e, quando aplicado ao estudo de QI-*cannabis*, é facilmente responsabilizado pelo suposto efeito de uso da *cannabis* por um longo tempo.⁶ Então, será que os curtidores podem simplesmente revidar e continuar fumando? Nem tanto, porque o efeito Flynn é apenas um fator de confusão em potencial, e não um elemento comprovado. O que está além de qualquer dúvida, porém, é a vulnerabilidade dos estudos de correlação em referência aos fatores de confusão – e a necessidade de continuar procurando-os mesmo quando temos a resposta “certa”. Isso é especialmente importante em estudos de fontes de risco comuns, mas controversas, como o fumo passivo, que são eles mesmos fatores de confusão em outras pesquisas.⁷

Tudo isso poderia dar a impressão de que as correlações são coisas traiçoeiras, que espalham armadilhas para os incautos. Todavia, há alguns sinais de alerta que sempre deveriam fazer soar um alarme. O primeiro é se os dados brutos foram agrupados para dar a impressão de que a coisa é mais bem-arranjada do que realmente é. Um jeito óbvio de fazer isso é pegar toda uma carga de medidas, tirar a média e correlacioná-las. O processo de tirar a média reduz todos os dados espalhados e confusos em pontos elegantes e bem-arrumados. O resultado pode ser um nível de correlação aparentemente muito mais expressivo – como perceberam muitos pesquisadores nas ciências mais “moles”. Num exemplo de livro-texto,⁸ a correlação entre nível educacional médio e renda, para homens com idades entre 25-54 anos em cada um dos estados americanos, foi determinada em +0,64, mostrando a importância de permanecer na escola. Mas quando a análise foi repetida usando dados do censo, a variação resultante fez baixar a correlação para +0,44.

Esse artifício de “limpeza de dados” é especialmente enganador quando a quantidade de dispersão viola um dos fundamentos da teoria da correlação simples: a de que a quantidade de variação permaneça constante. Por exemplo,

dados brutos podem vir de diferentes fontes de qualidade variável, ou pode simplesmente haver menos pontos de dados em alguns lugares que em outros. O resultado é mais incerteza e correlações potencialmente enganosas. Pesquisas de riscos assustadores para a saúde são particularmente vulneráveis a isso. Com frequência, há montes de pessoas com baixa exposição, mas relativamente poucas com exposição alta, aumentando a incerteza e o nível de dispersão à medida que o nível de exposição aumenta.

A dispersão também pode surgir das próprias variáveis. Talvez haja algum fator desconhecido em ação, ou talvez uma das variáveis simplesmente não tenha uma variância bem-definida; como vimos, há um bocado disso na natureza. E é possível que diversos desses efeitos ocorram simultaneamente. Seja lá o que for, a conclusão é que maneiras simples de tentar mascarar o problema por meio de médias bonitinhas e bem-arrumadas podem contribuir para construir gráficos mais convincentes, porém as correlações e outras inferências resultantes podem ser irremediavelmente enganosas.

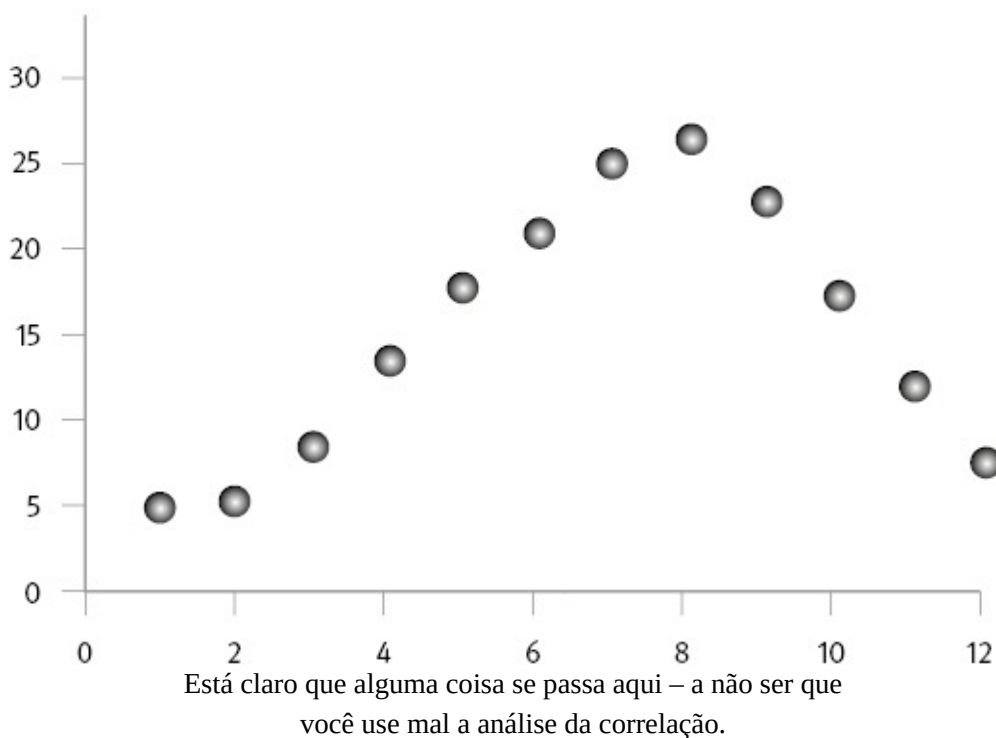


Dados correlacionados: a apresentação com versão elegante –
e o material bruto, confuso e incerto.

Advertências sobre como simples artimanhas com os dados podem solapar a confiabilidade das correlações vêm sendo dadas desde que elas foram usadas pelas primeiras vezes. De fato, o mesmo matemático que desenvolveu a teoria básica, Karl Pearson, advertiu os pesquisadores acerca de correlações baseadas em *proporções*, tais como X “por 1 000 pessoas” ou “por mês”. Elas são

frequentemente empregadas em áreas administrativas, bem como em pesquisa acadêmica, com o objetivo de colocar “tudo na mesma base”, mas tanto a pesquisa teórica quanto a empírica mostraram que os temores de Pearson são bem-fundamentados⁹ – o que é bastante preocupante, considerando a pletora de supostas “relações” construídas a partir de correlações baseadas em proporções.

Mais de meio século atrás, o celebrado estatístico Jerzy Neyman declarou que “correlações espúrias vêm arruinando a pesquisa estatística empírica desde tempos imemoriais”. A plausibilidade – ou a ausência dela –, além do velho ditado de que “correlação não é causalidade”, pode nos poupar de ler demais a partir de muito pouco. Mas não devemos esquecer que o reverso da moeda também é verdade: ausência de correlação não implica necessariamente ausência de uma relação genuína. Afinal, os “termos e condições” da teoria da correlação simples presumem que a relação seja linear, e há muitas que não o são. Deem uma olhada na figura a seguir.



À primeira vista, aí parece haver algum tipo de relação – mas a teoria da correlação simples nos diz que não há: o coeficiente de correlação é de apenas

0,36, e tem um valor p irremediavelmente não significativo de 0,25. Mas esses dois números na realidade só nos dizem duas coisas. Primeiro, se existe uma relação, ela não é uma simples linha reta, o que você pode ter sacado com uma olhadela no gráfico. Então o valor p diz que as chances de obter algo tão pobre quanto uma linha reta só por mero acaso são bastante altas – outra informação inútil. Mesmo assim, se ignorarmos – como faz um número exagerado de usuários de métodos estatísticos – as limitações da teoria da correlação simples (se é que algum dia as conhecemos), e interpretarmos erroneamente o alto valor p como “as chances de que o resultado seja mero acaso”, o assunto está acabado: aqui não acontece absolutamente nada. O que desafia o senso comum – e assegura que deixemos de ver uma informação-chave sobre quando visitar o Japão, pois os pontos mostram a ligação entre o mês e a temperatura típica em Tóquio, o que é tanto real quanto significativo – em cada sentido dessa palavra tão abusada.¹⁰

Conclusão

Correlações são como coincidências: nós as levaríamos bem menos a sério se tivéssemos mais consciência de como é fácil encontrá-las. Existem métodos poderosos para medir a correlação, mas eles se revelam enganosos se insistimos em que “deve haver alguma coisa nesse padrão”.

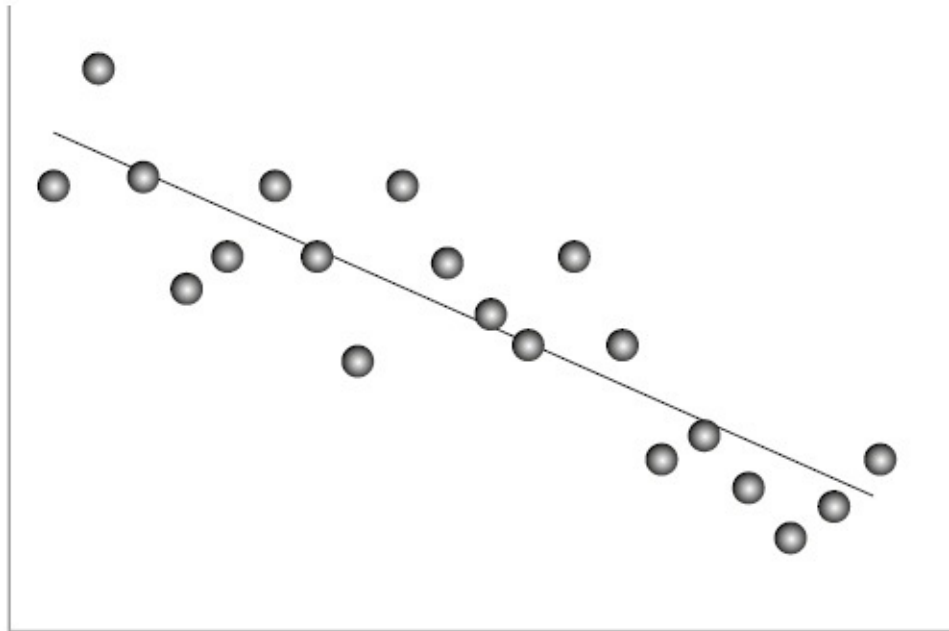
32. Temos de traçar a linha em algum lugar

NINGUÉM LANÇOU mão de forma mais impressionante das leis da física que a agência espacial americana, a Nasa. Em janeiro de 2006, ela disparou um objeto do tamanho de um piano de cauda na direção de um alvo a 4,5 bilhões de quilômetros de distância, movendo-se cerca de 50 mil quilômetros por hora. Nove anos depois, a sonda New Horizons passou zunindo por Plutão 72 segundos antes do horário programado, num encontro equivalente a acertar um *hole-in-one*¹ a uma distância de trinta quilômetros. A Nasa pode realizar esses feitos porque seus cientistas e engenheiros são muito espertos e realmente têm muito pouco com que se preocupar: apenas com o vácuo espacial entre eles e o alvo, os planejadores da missão podem sempre se virar usando a lei da gravidade e mais alguns truques para fazer previsões de impressionante confiabilidade. Eles conseguem declarar com confiança quase plena que, se lançarem com êxito numa data específica a tal e tal velocidade, em tal e tal trajetória, acabarão naquele ponto na data marcada.

De volta ao planeta Terra, as coisas não são tão simples, mas a mesma pergunta surge numa miríade de contextos: se ocorrer isto, o que acontece em seguida? Se a prevalência de gases do efeito estufa continuar aumentando, o que acontecerá com as temperaturas globais? Se cobrarmos mais pelo produto, qual será o impacto nas vendas? Se isto, então o quê?

Acontece que o método mais usado para descobrir foi inventado com propósitos astronômicos mais de duzentos anos atrás. O polímata alemão Carl Gauss – ele mesmo, famoso pela curva do sino – parece tê-lo usado para ajudar a (re)descobrir o primeiro asteroide conhecido, Ceres, em 1801. É o chamado método dos mínimos quadrados, ou, termo somente um pouquinho menos opaco, regressão linear. Em essência, ele simplesmente insere uma linha reta através de dados espalhados, mas não é uma linha reta qualquer; o método encontra a reta

que melhor se encaixa. A definição exata de “melhor” aqui é um pouco técnica,¹ mas em essência significa que ela representa uma tarefa matematicamente precisa do que você faria se lhe pedissem que inserisse uma reta mais próxima do máximo possível de dados:



A regressão linear encontra a “melhor” reta através de dados espalhados – até certo ponto.

Armados com essa “reta de regressão” mais bem-encaixada extraída dos nossos dados (qualquer planilha é capaz de fazer isso), todo tipo de coisa se torna possível. Podemos: usar a reta para preencher lacunas nos dados; utilizar a inclinação da reta para avaliar o impacto da mudança de uma variável sobre outra; ver quando e onde uma ou outra variável se torna zero; empregar a reta de regressão para ir *além* dos nossos dados. Imagine: poderíamos ter dados do mercado financeiro com preços de ações em diferentes momentos, recorrer à regressão linear para encaixar a melhor reta e então *prever* qual seria o preço amanhã, na próxima semana ou com meses de antecedência. E aí ficaríamos *ricos*.

Se você chegou até aqui no livro, deve ter sacado que deve haver algo de errado. O que você talvez não tenha percebido é como tantas pessoas inteligentes

não sacaram isso. O mais básico dos problemas em recorrer à regressão linear para achar relações entre dados é o mesmo com o qual tropeçamos na correlação: a simples ideia de haver alguma relação. Imputar uma causa a partir da correlação é arriscado mesmo quando é feito com cuidado. Quando é feito de maneira descuidada, os resultados, na melhor das hipóteses, são risíveis. Alguns cliques numa planilha permitem que a regressão linear deduza a lei de Nicolas Cage em toda sua sutileza matemática:

$$\text{Nº de afogamentos} = 5,8 \times \text{nº de filmes de Nicolas Cage} + 87$$

Adicionamos como cobertura do bolo um coeficiente de correlação altamente expressivo (+0,67), acrescentando como enfeite final a cereja de a correlação ser estatisticamente significativa ($p = 0,025$). Levada a sério, essa equação de regressão nos diz que cada novo filme estrelado por Nicolas Cage causa mais seis mortes por afogamento. No entanto, ninguém levaria isso a sério; a lei é um patente absurdo, porque... é, ponto final. E aí reside o problema da análise de regressão: ela não diz nada sobre se chegou a fazer algum sentido experimentá-la. Ainda estamos à espera de um programa de cálculo que identifique tentativas desaconselháveis de achar relações em dados e inserir a reta mais apropriada junto com a mensagem: “Você só está de brincadeira, certo?”

Um pouquinho mais alto na escala de sofisticação é presumir que tudo bem encaixar uma reta nos dados que realmente reflita algo mais complexo. Novamente, não há sentido em procurar conselhos em algum programa de computador. Como Igor, o fiel assistente do dr. Frankenstein, esse programa fará automaticamente qualquer coisa que lhe peçamos, não importa quão pavoroso seja o resultado. Mesmo que os pontos dos dados sigam o contorno de uma banana caramelada, a regressão linear introduzirá a reta mais bem-encaixada entre eles.

E até nos fará sucumbir ao desejo de agir como deuses e prever o futuro. E por que não haveríamos de fazê-lo? Afinal, se podemos usar a regressão linear para mostrar, por exemplo, como as vendas de um produto variam com os gastos de publicidade, por que não recorrermos ao mesmo expediente para prever as

vendas ao longo do tempo? Não há nenhuma razão que impeça – exceto que o tempo não é só mais uma variável. Ele tem o péssimo hábito de ligar as coisas. Por sua vez, isso suscita o velho problema: não ler os “termos e condições” do kit matemático que estamos usando.

Enterrada no meio dos “termos e condições” da regressão linear está a exigência de não haver padrão nos erros cometidos pela reta “mais bem-encaixada” quando ela passa através dos pontos que representam dados. Como sempre, isso parece chato e complicado, mas, como ocorre tantas vezes, também é crucial, pois é o que pode aparecer nos dados cobrindo um intervalo de tempo. Tudo, desde ciclos nos negócios e efeitos sazonais até simples impulsos, é capaz de gerar vínculos entre pontos de dados, e a “autocorrelação” resultante talvez faça troça de qualquer previsão baseada em regressão. Felizmente, há todo um arsenal de técnicas para lidar com isso como parte de uma enorme e fascinante disciplina chamada análise de séries temporais. A má notícia é que ela exige conhecimento especializado para manejá-la. Pior ainda, mesmo aqueles que têm esse conhecimento ainda podem acabar, e acabam, metidos em encrenca.

Vamos tomar o relato do Google Flu Trends (Tendências de Gripe do Google, GFT), que provocou um rebuliço, com sua suposta capacidade de emitir alertas precoces acerca de surtos de gripe letal. Num artigo publicado na revista *Nature* em 2009, analistas de dados da empresa de tecnologia e peritos dos Centros de Controle de Doenças (CDC, o Centers for Disease Control) dos Estados Unidos alegavam ter identificado a estação de gripes de 2007-08 uma semana ou duas antes da rede de detecção dos CDC.² Eles o fizeram varrendo anos de dados armazenados no colossal arquivo histórico do Google, caçando correlações entre surtos de gripe e termos digitados no mecanismo de busca da empresa. Em vez de tentar adivinhar que termos eram mais preditivos, a equipe entregou a tarefa aos computadores, que experimentaram a estupefaciente quantidade de 450 milhões de modelos. O melhor usava 45 termos de busca para produzir uma expressiva correlação 0,97 com surtos futuros.

Esse foi um impressionante exemplo de mastigação de dados. Por algum tempo o GFT parecia anunciar uma nova era, na qual imensos conjuntos de

dados e potência computacional insuflavam nova vida em esgotadas e antigas técnicas, como a regressão e a correlação. Todavia, os “termos e condições” eram poderosos como sempre, e logo impuseram sua autoridade. O algoritmo do GFT mal fora tirado da sua caixa quando falhou, perdendo completamente um surto de gripe em 2009 e forçando seus criadores a remendar o programa, o que não fez muita diferença: as previsões do GFT continuaram pouco melhores que os métodos tradicionais dos CDC e tinham o hábito de superestimar o tamanho dos surtos. Em 2014, uma equipe de analistas de dados da velha escola publicou uma contundente análise do desempenho do GFT, deixando claras suas inadequações. Estas incluíam falha em lidar com o conhecido problema das séries temporais, a autocorrelação.³ No ano seguinte, o Google fechou o site do GFT e ofereceu seus dados a qualquer pessoa que se julgasse capaz de se dar melhor. É inteiramente possível que haja um “sinal” útil enterrado ali em algum lugar; menos clara é a melhor maneira de extrair esse sinal – e se o esforço chega a valer a pena.

Mas há uma informação que não pode ser negada. Mesmo antes do lançamento do GFT, havia alegações de que os conjuntos colossais de dados significavam que não era mais necessário se afligir por causa dos “termos e condições” – nem sequer saber o que se devia fazer. Em vez disso, os dados podiam ser simplesmente jogados em massa dentro do computador, que compararia tudo com o resto até achar as melhores correlações possíveis. Não era preciso haver compreensão, modelos ou mesmo palpites; nas palavras de um comentarista de olhos arregalados: “Com dados suficientes, os números falam por si sós.”⁴ O vexame do GFT mostrou que, segundo um celebrado perito em dados, isso era “uma completa besteira, um total absurdo”.⁵ O fato é que esse reluzente novo campo dos “Big Data” está sujeito aos mesmos cansativos, embora cruciais, “termos e condições” dos Small Data – com armadilhas adicionais, em grande medida. Quem pensar em pegar uma pá digital e cavucar vastos conjuntos de dados deve ter isso em mente. Se até os gênios brilhantes do Google podem acabar com pouco mais do que ouro de tolo, pense só no que a garimpagem de dados pode fazer com você (ver Box a seguir).

Nada disso deteve a torcida organizada dos Big Data. Com uma crença evangélica no poder miraculoso de métodos tais como regressão, eles produziram grandes notícias no ramo dos grandes negócios. Em 2014, um levantamento global descobriu que aproximadamente três quartos das organizações terão investido em tecnologia de Big Data em 2016; o mercado já vale cerca de US\$ 125 bilhões.⁶ As principais prioridades são usar a tecnologia para “aprimorar a experiência do cliente” e “aperfeiçoar a eficiência do processo”. Todavia, já há sinais de sérios problemas pela frente. Pessoas dentro das indústrias já avisam que as empresas planejam minerar praticamente tudo e qualquer coisa em seus arquivos de dados – estratégia segura para achar ouro de tolo. No final, porém, os Big Data viverão ou morrerão no mundo empresarial de acordo com o critério secular: ele aumenta os lucros? Isso está longe de ser garantido. Uma das primeiras histórias sobre Big Data centrava-se num prêmio de US\$ 1 milhão oferecido pelo serviço de filmes on-line Netflix, em 2006, para quem conseguisse garimpar um modo melhor de predizer as avaliações sobre os filmes. Três anos depois, uma equipe embolsou o prêmio, mas a Netflix nunca pôs o algoritmo em funcionamento. Apesar de atender ao requerido aumento de 10% na performance, era incrivelmente complexo, e a empresa resistiu a pagar pelo upgrade em tecnologia necessário para obter benefício tão pequeno.⁷ À medida que os *data mining* entrarem num mundo mais amplo, eles enfrentarão encontros igualmente duros com a realidade. Diretores de vendas podem não conhecer os perigos da autocorrelação, mas sabem quando suas previsões de vendas baseadas em *data mining* estão furadas.

CUIDADO: DATA MINING EM ANDAMENTO

Data mining – garimpagem de dados – é um negócio global de US\$ 100 bilhões, e todo mundo, de multinacionais a enxovais de papai e mamãe, se digladia para usá-la. Então, por que tantos veteranos de análise de dados não chegam a se extasiar com a revolução dos Big Data? Depois de passar muito tempo tentando extrair informações de pequenos punhados de dados, eles deveriam se deleitar ao pôr as mãos em conjuntos de dados realmente colossais. Contudo, décadas “se contentando” com pouco lhes ensinaram algumas duras lições que se aplicam a todos os conjuntos de dados, grandes ou pequenos. Tomemos o problema do viés:

1 bilhão de pontos de dados de fontes seletivas são potencialmente mais capazes de levar ao erro que uma minúscula fração obtida de uma amostra adequadamente randomizada (por exemplo, quem são exatamente as pessoas que buscam remédios para gripe no Google e por quê?).

Ainda assim, uma vez que você consegue um conjunto de dados limpo, sem viés, é fácil criar um modelo de previsão a partir deles. Basta usar regressão e análise de regressão num computador para achar influências estatisticamente significativas, e então combiná-las para obter um encaixe perfeito nos dados. Ao contrário, é aí que está o desastre. Quando se deixa que um conjunto de dados “fale por si só” dessa maneira, ele faz jorrar absurdos. Sem nenhuma tentativa de extirpar correlações implausíveis, acaba-se confiando na “significância estatística” para avaliar a relevância. Lamentavelmente inadequada, na maioria das vezes ela pode se mostrar catastrófica. Emparelhar apenas dez variáveis uma com a outra enquanto se caçam correlações “reais” significa um risco de 90% de encontrar pelo menos uma que seja estatisticamente significativa por puro acaso. A garimpagem de dados muitas vezes envolve uma quantidade muito maior de variáveis. Um modo de cortar o risco é ajustar o padrão para a significância. Isso ajuda, mas surge um fenômeno muito estranho: o paradoxo de Jeffreys-Lindley. Há muito tempo conhecido entre os estatísticos, ele implica que, quanto maior o conjunto de dados, *menos* efetivos são os testes de significância para identificar achados fortuitos. Outra surpresa desagradável aguarda aqueles que pensam que os algoritmos de previsão devem incluir idealmente o máximo possível de variáveis. Ao mesmo tempo que fornecem uma combinação expressiva com dados já arquivados, esses algoritmos podem falhar terrivelmente quando ganham vida. O problema reside no chamado dilema do viés de variância. Mais variáveis fornecem previsões mais acuradas, menos enviesadas, para combinar bem com velhos dados, mas sofrem com dados novos. Como cada variável tem sua própria incerteza, a turbidez (“variância”) da previsão também aumenta. Cabe haver uma compensação: apenas variáveis suficientes para fazer um bom serviço, mas não tantas que tornem as previsões irremediavelmente vagas.

Todos os desafios podem ser enfrentados – se forem reconhecidos desde o início. Contrariamente ao que podem alegar alguns, quando se trata de garimpagem de dados, o tamanho não é tudo.

Para os ansiosos por usar o poder dos *data mining*, as preocupações sugeridas pelos analistas da velha escola são vistas como reacionárias e excessivas. Afinal, os cientistas não vêm usando técnicas como a regressão em pesquisa há décadas, sem nenhum problema óbvio? Embora os cientistas venham realmente usando essas técnicas, a confiabilidade daquilo que descobriram é pouco certa. Seria errado pensar que os cientistas sempre manejaram as ferramentas dos *data mining* com cuidado. Um caso ilustrativo é o

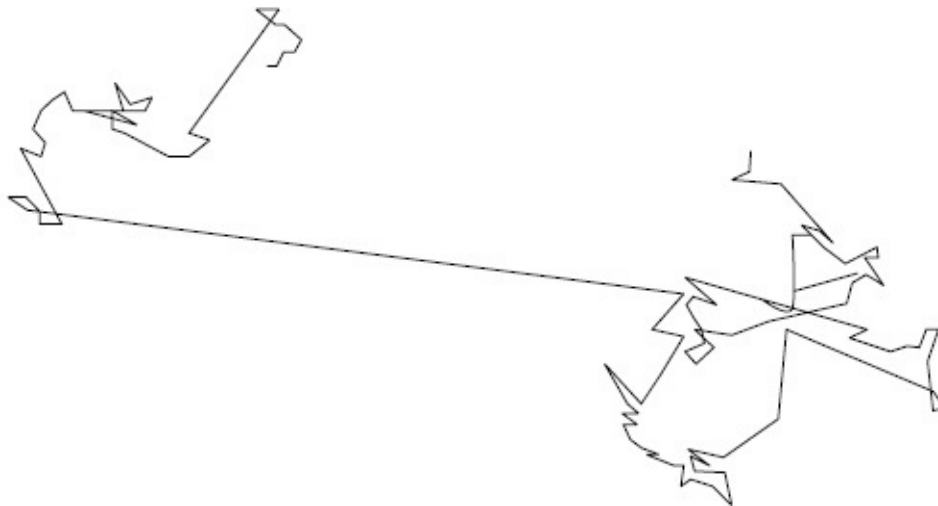
salutar relato da febre das leis de potência. Durante a década de 1980, as mais importantes revistas científicas começaram a receber artigos alegando que fenômenos desde movimentos de mercado até provisões de alimentos de formigas seguem as chamadas leis de potência da forma:

$$\text{Algo interessante} = k \times (\text{algo mensurável})^N$$

Os artigos se concentravam em achar o valor da potência N , pois esta levava a uma série de teorias e ideias interessantes. Para descobrir qual era esse valor, os pesquisadores usavam um artifício simples, que lhes permitia aplicar o método da regressão linear a todos os tipos de conjuntos.⁸ Os valores resultantes de N geraram outra onda de artigos dedicados a explicar como e por que essas leis de potência existiam. Em meados dos anos 1990, um dos principais expoentes da lei de potência sentiu-se encorajado a redigir um livro de popularização sobre tudo aquilo, com o modesto título *How Nature Works*.⁹ Mesmo na época, essas alegações provocaram cenhos franzidos, mas levou um tempo longo demais para que as restrições se transformassem em críticas. O porquê é uma questão interessante na sociologia da ciência, já que estava claro desde o início que alguns pesquisadores cometiam uma extrema violência com vários dos “termos e condições” da regressão linear.¹⁰ Na sua determinação de achar N , arriscaram-se a chegar a conclusões absurdamente não confiáveis. Algumas das que mais chamavam a atenção eram os argumentos de que as leis de potência serviam de base para uma estonteante variedade de organismos. Dos anos 1980 em diante, pesquisadores reivindicaram ter descoberto que os modelos de busca de alimentos e caça de muitas criaturas seguem padrões conhecidos como voos de Lévy. Uma vez mapeados, estes parecem aglomerados aleatórios de pequenas excursões seguidas de outras maiores, mais raras – cujas proporções relativas seguem uma lei de potência.

Várias explicações foram apresentadas, todas argumentando que a mistura de etapas curtas e longas era de algum modo “ideal” para a busca de alimento. E parecia que era explorada por uma grande quantidade de organismos, desde abelhas e albatrozes, no ar, até oceânicos, plânctons, focas e mesmo tripulações humanas de barcos pesqueiros. Mas essa “evidência” se pautava amplamente em

regressão linear – que pode ir para o brejo quando alimentada com dados acerca de tais fenômenos. Em 2005, o ecólogo e matemático Andrew Edwards, agora na Fisheries and Oceans Canada, começou a investigar a base dessas alegações, e analisou-as novamente, usando técnicas melhores, capazes de lidar com a natureza teimosa das leis de potência. Ele descobriu que, de dezessete supostas descobertas publicadas, nenhuma resistia a um exame detalhado.¹¹ Desde então, os pesquisadores passaram a rever todo o tema usando os métodos mais avançados, e – pelo menos no caso dos albatrozes – descobriram que as alegações originais podiam estar certas, embora pelos motivos errados.¹²



“É um pássaro? É um rabisco? Não, é um voo de Lévy
– e um alerta sobre forçar os dados longe demais.

Essa é uma boa nova para os ecólogos, mas deixa sem resposta a pergunta: exatamente quanta bobagem baseada em regressão ainda está por aí, sem ser reconhecida? A menos que alguém resolva voltar e checar, provavelmente nunca saberemos. Com tanta pesquisa e tantas reputações que agora se assentam sobre resultados baseados em regressão, o pesquisador que resolver descobrir deverá ter muita coragem.

Conclusão

Todos os conjuntos de dados contêm padrões, mas a maioria é ilusória. Encontrar a reta que “melhor se encaixa” não muda isso. Apesar da badalação, os Big Data continuam vulneráveis

Gigo – de Garbage In, Garbage Out, ou “entra lixo, sai lixo”. Acrescentem-se a isso os muitas vezes ignorados “termos e condições” dos métodos de *data mining*, e você tem na mão uma técnica do século XXI para gerar absurdos.

¹ *Hole-in-one*: quando se acerta um buraco, no golfe, com apenas uma tacada. (N.T.)

33. Jogar com os mercados *não* é uma ciência precisa

QUANDO AS PESSOAS DESCOBREM que vários trilhões de dólares que julgavam estar bem guardados sumiram, elas tendem a exigir respostas. Em 2007, os escravos assalariados do mundo tinham cerca de US\$ 27 trilhões em seus fundos de pensão. A maioria contava com o dinheiro para lhes dar uma qualidade de vida razoável após décadas batendo ponto. Muitos já viviam de seus modestos pecúlios, contando com os mercados de ações para continuar melhorando com os rendimentos e dividendos. Então a crise financeira atacou, o mercado de ações desmoronou, e o valor dos fundos de pensão no mundo caiu para US\$ 3,5 trilhões.¹

Na busca dos culpados, os holofotes recaíram imediatamente sobre Wall Street, a City de Londres e os bancos de investimentos em todo lugar. E então focalizaram os habitantes desses templos da cobiça, com seus Porsches e a avidez por bônus, o cabelo lambuzado de gel e esquemas de enriquecimento rápido. Mas logo se voltou para aqueles que tinham concebido esses esquemas: os “cientistas exatos” metidos a gênios, com seus óculos sem aro e diplomas de doutorado. Enquanto outros viveram um tempo sob os holofotes, os analistas quantitativos – *quants* – ali permaneceram. Todos foram acusados de criar um arsenal de armas de destruição monetária que teria levado a economia global à beira do desastre. Vários relatos iniciais da crise se concentravam em como os *quants* vinham se envolvendo na “engenharia financeira”, criando o que chamaram de “derivativos” com nomes bizarros como CDS – Credit Default Swaps – e Bermuda Swaptions. Também inventaram as chamadas “seguridades” – obviamente uma piada para quem estava por dentro – de nomes estranhos como ABS-CDOs. Mas eram os mecanismos internos dessas armas de destruição monetária que provocavam reais convulsões de horror. Eles eram embalados por modelos matemáticos medonhos e complexos, que apresentavam

uma impressionante semelhança com a física teórica. A conclusão parecia clara: o sistema financeiro global havia sido capturado pelos cientistas malucos.

Esse cenário assustador desde então tem sido questionado em diversas frentes. Primeiro, derivativos não são uma coisa nova: a ideia básica de uma promessa financeira respaldada em um contrato para enfrentar uma possível inadimplência remonta a milênios.² Segundo, longe de serem apenas esquemas de enriquecimento rápido, havia muito eram vistos como essenciais para o comércio, trazendo pelo menos um bocadinho de confiança acerca de um futuro incerto. Mas a ideia de que esses bancos de investimentos estavam cada vez mais lotados de físicos sonhando com “instrumentos” cada vez mais malucos também é um mito. Na realidade, não há escassez de gente conhecedora de matemática no meio das finanças, mas, por outro lado, tem havido apenas um número relativamente modesto de físicos nesse ambiente. Essa é uma distinção crucial, primeiro porque os físicos estão entre os críticos mais preeminentes do uso de técnicas financeiras complexas,³ e também porque eles conhecem o segredinho sujo sobre sua disciplina.

A verdade acerca da física, e sua importância central para compreender a crise financeira, é o foco de uma das análises mais perspicazes sobre o desastre, publicada em 2010. Ela traz o curioso título de “Aviso: inveja da física pode ser prejudicial para sua riqueza”,⁴ e as credenciais de seus autores não são menos intrigantes: Andrew Lo, distinto professor de finanças na Escola Sloan de Administração do Massachusetts Institute of Technology (MIT), e Mark Mueller, físico do MIT que largara a área na década de 1990 para se tornar parte da mesma comunidade agora acusada de sucatear a economia global: a POW, ou Physicists on Wall Street (Físicos em Wall Street). Juntos, ambos examinaram a ideia de que a crise financeira era resultado da síndrome que constava do título do seu estudo.

Apesar do nome bem-humorado, a inveja da física é uma síndrome genuína – e não é de admirar. De todas as ciências, nenhuma conquistou mais sucesso, credibilidade e renome que a física. Suas descobertas sustentam o mundo moderno e servem de informação para as nossas noções de realidade. Seus

maiores praticantes são sinônimos para a palavra gênio, suas grandiosas teorias são saudadas como as maiores conquistas do intelecto humano. Quem não gostaria de ter um pedacinho disso? Depois da Segunda Guerra Mundial, enquanto os físicos se aqueciam ao sol da gratidão global por terem ajudado a derrotar o mal, estudiosos de outras áreas começaram a se perguntar o que o “caminho da física” podia fazer por eles. Talvez eles também conseguissem identificar leis fundamentais e as usassem para modelar a realidade e moldar o futuro em benefício da humanidade?

Entre eles estava Paul Samuelson, genial estudante de economia que entrou na Universidade de Chicago aos dezesseis anos e completou o doutorado em Harvard aos 22. E não foi um doutorado qualquer: publicado em 1947 sob o título de *Fundamentos da análise econômica*, tornou-se precisamente isso – e levou Samuelson a se tornar o primeiro americano a ganhar o Prêmio Nobel de Economia. Aquele era o reconhecimento – conforme a citação do prêmio, em 1970 – de seu “trabalho científico” e de seu sucesso para “elevar o nível da análise na ciência econômica”. Samuelson foi o catalisador de uma virada radical em economia e finanças, afastando-as das baboseiras e do raciocínio “de senso comum”, assestando-as em direção à abordagem matemática e com princípios que tão bem haviam servido aos físicos durante tanto tempo. Na verdade, Samuelson seguia os passos de outros que julgavam que a física tinha muito para ensinar aos economistas. No começo do século XX, um matemático francês chamado Louis Bachelier havia aplicado a teoria da probabilidade aos preços das ações, encontrando evidência de que elas pareciam se comportar como se estivessem sob a influência de forças aleatórias. Cinco anos depois, Einstein desenvolvia uma explicação semelhante para os movimentos bruscos de partículas microscópicas – usando os resultados para inferir a realidade dos átomos. O próprio mentor de Samuelson, Edwin Wilson, era polímata e protegido do brilhante físico americano Josiah Willard Gibbs.

Durante as décadas de 1950 e 1960, economia e finanças tornaram-se mais parecidas com a física; suas publicações especializadas eram cada vez mais impenetráveis para aqueles não familiarizados com as ferramentas da física,

como a álgebra linear e o cálculo integral. Todavia, como Samuelson, ainda com vinte e poucos anos, viu, as semelhanças podem ser perigosamente enganosas. Como iniciado no segredinho sujo da física, ele sabia que, apesar de toda a sua aparente complexidade e sofisticação, a física tem êxito porque se concentra nos problemas que são essencialmente *simples*. Isso parece risível para quem abandona a física quando começam as aulas focalizadas em trajetórias parabólicas de bolas lançadas de precipícios. Contudo, ainda que as equações quadráticas pareçam abstrusas, elas funcionam porque o problema se torna *tão* simples que as equações oferecem uma perspectiva proveitosa. Adicione-se algum realismo – como a resistência do ar –, e a matemática logo se torna alucinante.⁵ Economia e finanças não compartilham com a física esse nível de complexidade – elas são *muito mais* complexas, repletas de fenômenos que desafiam o uso da matemática. A análise teórica de Einstein acerca do comportamento dos átomos foi um avanço fundamental porque suas características-chave são verdadeiras para sempre. No entanto, de modo paradoxal, isso também tornou a tarefa de Einstein muito mais simples. Imagine se, em vez disso, os átomos às vezes resolvessem correr numa direção, ou responder de forma diferente a forças idênticas. O problema seria muito mais difícil, complicado e menos “fundamental” de se enfrentar. Mas também seria surpreendentemente relevante para o comportamento dos “átomos” do mercado de ações: os investidores humanos.

Samuelson achava que a abordagem física na economia e nas finanças era proveitosa, mas só até certo ponto. Compreendeu que, com toda sua grandiosa teorização, os físicos na verdade tinham vida fácil. Muitos dos seus maiores triunfos baseiam-se na exploração da simetria – em essência, a capacidade de qualquer coisa ser alterada de alguma maneira, e no entanto permanecer inalterada. Poder assumir essa constância no espaço e no tempo simplifica enormemente as teorias, desde partículas subatômicas até o Universo.⁶ Os economistas não têm nada que se compare a isso: no seu universo, a única constante é a mudança. Na verdade, como apontaram Lo e Mueller, a situação é ainda pior. Além de não poderem contar com a constância dos “átomos” em suas

teorias, os economistas não conseguem sequer saber se ou quando suas teorias chegam a se aplicar.

A tragédia da economia nos anos pós-guerra foi que o ofuscante sucesso da física cegou muita gente para o seu segredinho sujo. Inúmeros economistas passaram a encarar o uso extensivo da matemática pelos físicos como sinal de sofisticação, e não como sintoma de simplificação. Os físicos devem ser invejados pelo fato de receberem reverências por explorar mundos que toleram simplificação a ponto de se aplicar a eles a matemática – e ainda restar alguma coisa que valha a pena dizer. Eles podem jogar fora a água do banho da natureza impunemente, já que achar um simples pato de borracha basta. Os economistas, em contraste, querem saber se os bebês ficam mais felizes espirrando água em banhos mais caros. Eles também podem jogar a água do banho fora, mas arriscam-se a jogar o bebê junto, e, de um ou de outro jeito, acabam fazendo sacadas que não têm sentido. Em vez de invejar os físicos pela simplicidade do que fazem, os economistas os invejavam pelo seu sucesso ao fazê-lo. Mas pelo menos reconheceram a incerteza muito maior dos problemas em economia e nas finanças. Fazer previsões, investir, projetar derivativos, tudo isso exigia adotar uma visão sobre o futuro incerto. Em sua busca de “matematizar” a disciplina, os economistas voltaram-se então para a teoria da probabilidade. Mas estava claro que a versão mais básica não era suficiente. A economia lida com situações muito mais complexas que lançamentos de moedas ou dados, em que as probabilidades são fixadas e óbvias. Os mercados financeiros são resultado de múltiplas influências, todas elas sujeitas a incertezas. Assim, compreendê-las exigia dos economistas passar para o patamar seguinte da teoria da probabilidade, que captava o efeito de influências aleatórias múltiplas. Como vimos, isso levou ao pesado uso da distribuição normal, cuja elegância e a potência em lidar com a atrapalhada incerteza da vida real haviam sido reconhecidas mais de um século antes.

No entanto, como também vimos, toda técnica para lidar com a incerteza chega com termos e condições – alguns dos quais eram claramente violados nas situações em que os economistas usavam a técnica. Evidência dessas falhas

podia ser encontrada em dados empíricos, mas durante anos os estudiosos que as apontavam viam sua pesquisa rejeitada pelas revistas mais influentes de economia e finanças.⁷ E havia um problema ainda mais fundamental, que simplesmente não podia ser matematizado. Pairando ameaçador sobre as incertezas dos fenômenos econômicos e os modelos dessas incertezas havia algo muito maior: a incerteza em relação aos próprios modelos. Como Lo e Mueller ressaltaram, isso coloca a economia num território onde até mesmo os físicos caminham com temor, onde modelos nos quais se confia caem por terra e precisam ser substituídos. Os físicos depararam com essas situações muitas vezes no correr dos séculos. As leis do movimento de Galileu ruíram quando forçadas longe demais, e deram lugar à teoria da relatividade especial de Einstein; a visão de espaço, tempo e gravidade de Newton tornou-se incorporada à teoria da relatividade geral de Einstein; a visão de átomos como minúsculos sistemas solares deu lugar às nuvens probabilísticas da mecânica quântica. Os físicos se recompuseram, aprenderam os limites dos velhos modelos e os empregaram a fim de selecionar o melhor para cada tarefa.

Mas os economistas podem acordar amanhã e encontrar o equivalente a uma suspensão da lei da gravidade. Ontem, tudo estava bem; hoje, a Rússia, por exemplo, dá o calote na sua dívida estatal, fazendo com que alguns mercados se despedacem como areia, experimentando uma lei inversa da raiz quadrada da gravidade. Ao mesmo tempo, outros decolam como se a gravidade tivesse sido desligada. Os modelos padronizados não funcionam mais, e, embora possam voltar a atuar com o tempo, ninguém é capaz de dizer quando.

Em face desse modelo de incerteza, a idílica matemática é impotente. Só uma coisa pode impedir o desastre: o uso judicioso do dispositivo mais complexo no Universo conhecido – o cérebro humano. Os brinquedinhos reluzentes de equações diferenciais parciais e cálculo Itô precisam dar lugar à solidez da experiência, do julgamento e da determinação.

A crise financeira teve muitas causas, políticas, regulatórias e psicológicas entre elas. Contudo, todas têm suas raízes no mesmo fenômeno: seres humanos tentando lidar com a incerteza. Os execrados “cientistas exatos” lidavam com ela

empregando modelos cada vez mais complexos, na esperança de que o diabo estivesse nos detalhes. Outros lidavam com ela tentando ganhar o máximo de dinheiro, de modo a não haver incerteza quanto ao seu próprio futuro. Mas todos eram superados em número pelos administradores e diretores executivos, reguladores e legisladores que de boa vontade caíram sob o feitiço da inveja da física e da crença de que os artifícios que revelam os segredos cósmicos seguramente devem funcionar nas finanças. Mesmo agora, ainda não está claro quantos deles finalmente acordaram para o fato de que lidar com as incertezas no mundo financeiro exige uma perícia muito além do que meramente matemática.⁸

A maioria dos físicos se orgulha de fazer parte da mais bem-sucedida das disciplinas científicas, ao mesmo tempo mantendo-se cômicos das limitações do seu *modus operandi*. Talvez um número maior deles devesse acompanhar gente como Lo e Mueller, deixando os outros penetrarem no segredinho sujo da física – antes que provoquem outro enorme buraco na economia global.

Conclusão

A crise financeira foi uma demonstração no valor de muitos trilhões de dólares dos perigos da inveja da física. Ao mesmo tempo que a matemática sofisticada do tipo produzido pelos físicos pode ser necessária nas finanças, com certeza ela não é suficiente. A física pode contar com as incertezas, enquanto as finanças envolvem não só uma hoste de incertezas, mas também a incerteza sobre essas incertezas.

34. Cuidado com geeks criando modelos

SE NÃO SE PODE confiar nos melhores e mais brilhantes pesquisadores em finanças para manter nosso dinheiro seguro, o que o comum dos mortais deve fazer? A primeira lição da crise financeira parece bastante clara: seja cético diante de quem alega que domou a incerteza. Isso é mais fácil de falar que de fazer, pois em geral essas pessoas têm doutorados, vêm com modelos de complexidade bizantina e até evidências brutas de seu sucesso ao longo de anos. Em física, tudo isso de fato seria impressionante, como seria qualquer evidência de progresso com valor duradouro. Mas aqui não se trata da física, com suas leis fundamentais e constantes universais. Isso são finanças, área onde os modelos às vezes são apenas simulacros de certeza. Eles podem efetivamente funcionar, mas só enquanto seus termos e suas condições valerem, e ninguém sabe o que isso significa. Talvez décadas, talvez dias.

A tentação de ignorar tudo e mergulhar de cara se cristaliza nas fortunas – em todos os sentidos – dos fundos de cobertura, ou fundos de hedge. Essas dissimuladas instituições são famosas por contratar os mais espertos geeks^m para conceber estratégias de *hedging* (“cobertura”) que deem o máximo de retorno com um mínimo de risco. Também são renomadas pelos chamados esquemas dois e vinte, em que os clientes pagam aos fundos 2% dos ativos pelo privilégio de eventualmente se beneficiar do brilhantismo coletivo da administração do fundo, mais 20% de qualquer lucro que esse brilhantismo efetivamente obtenha. E se a imprensa financeira pode servir de guia, esse é um preço que vale a pena pagar: os fundos de cobertura rotineiramente ganham as manchetes pelo seu talento em localizar oportunidades e evitar calamidades. Mas claro que a imprensa não serve absolutamente de guia: ela focaliza performers excepcionais que depois “regridem à média” – e os registros mostram que essa performance

média não é melhor que aquela obtida pelas estratégias de investimento convencionais, uma vez subtraídas as pesadas taxas administrativas.¹

Em suma, investir no fundo de cobertura típico é comprar uma prova cara de que até mesmo os mais complexos modelos financeiros estão sujeitos à incerteza. O real brilhantismo dos fundos de cobertura reside no esquema do negócio, que garante uma remuneração durante o tempo em que conseguem convencer os investidores a manter a fé nas estratégias – que pode ser mais longo que o próprio tempo de funcionamento das estratégias.

Felizmente, investir em fundos de cobertura é em grande parte um jogo que só os investidores ricos podem jogar. O resto de nós provavelmente acaba com investimentos administrados segundo estratégias ganhadoras do Prêmio Nobel. Por infortúnio, isso não é motivo para comemorar, pois essas estratégias surgiram de uma das mais egrégias tentativas de reduzir as complexidades das finanças para as simplicidades da física. Na Universidade de Chicago, no começo dos anos 1950, um estudante de economia de vinte e poucos anos se propôs a fazer pelas carteiras de investimentos o que Newton tinha feito pelos corpos em movimento. O resultado daria a Harry Markowitz uma participação no Prêmio Nobel de Economia de 1990. Voltando aos anos 1950, os conselhos dos especialistas para investir eram tão simples quanto ridículos: encontrar uma ação de desempenho top e botar todo o dinheiro nela. Markowitz sabia que isso fazia pouco sentido, assim como a maioria dos investidores. Estes haviam percebido que tinha muito mais cabimento possuir um mix “diversificado” de ativos no portfólio, para diluir o risco de perder tudo num desastre. Mas qualquer um que se sentasse para criar esse portfólio logo tropeçava num problema: qual devia ser o mix? Metade em ações vigorosas e agitadas, metade em letras do Tesouro, mornas mas seguras? Ou será que isso é seguro demais? Que tal uma divisão 80/20 entre ações e letras fixas... ou talvez 60/30/10, com os 10% em coisas que têm liquidez imediata? Markowitz percebeu que essas perguntas caíam no âmbito de um ramo da matemática aplicada chamado otimização restrita. O que ele precisava fazer era encontrar um mix ideal de ativos que minimizasse o risco, ao mesmo tempo dando um retorno decente.

As equações que ele anotou tornaram-se o alicerce para o que hoje se chama teoria moderna do portfólio (TMP). E, à primeira vista, elas conseguem algo milagroso. Você as alimenta com dados históricos sobre os ativos no seu portfólio, e elas revelam o mix ideal de ativos que você deve manter. No entanto, apesar do nome, a TMP não é uma teoria: é um modelo, e como tal está repleta de “termos e condições”, e premissas que variam de questionáveis até pura e simplesmente erradas.

Tomemos o conceito de “risco”. A maioria das pessoas acha que minimizar o risco significa minimizar as chances de sofrer uma perda longa e contínua. No entanto, ao buscar uma maneira de modelar matematicamente o risco, Markowitz pegou o conceito estatístico de variância – uma medida das oscilações de um papel financeiro em torno do seu valor médio. Isso parece estranhamente confuso, mas Markowitz se apegou à variância porque lhe permitia explorar um belo teorema em probabilidade que destravou todo o problema da otimização. Em poucas palavras, o teorema fornecia a ligação entre o risco total de um portfólio e o risco de cada papel nele contido, e como se correlacionavam entre si. Ou pelo menos era isso que acontecia se você acreditasse, como Markowitz, que a variância é uma boa medida de “risco”. Em caso positivo, você era devidamente recompensado com uma descrição matemática das características básicas de investir: os riscos e retornos dos papéis, e até a forma como cada um se movia com ou contra o outro.

As equações de Markowitz confirmavam a ideia de senso comum de que faz sentido ter um mix de ativos, mas iam adiante, mostrando precisamente como era o aspecto de uma “boa” diversificação. Esta exigia papéis com correlações mútuas baixas ou de preferência negativas, o que também fazia sentido, pois quando um valor cai, outros sobem para compensar as perdas. As equações também continham algumas surpresas – tais como os benefícios de incluir papéis mais arriscados. Se estes estivessem anticorrelacionados com os outros, poderiam efetivamente *reduzir* o risco geral do portfólio.

Tudo que um investidor, mesmo que fosse novato, tinha a fazer para explorar o poder da TMP era examinar a performance passada de alguns papéis e

estabelecer seus retornos, o grau de correlação e seu risco (medido pela variância de seus retornos).² Inseridos nas equações de Markowitz, os dados seriam magicamente convertidos nas divisões percentuais entre os papéis necessárias para criar o portfólio ideal, diversificado para minimizar o risco e com um retorno decente.

No entanto, como inúmeros investidores descobririam ao longo dos anos, à parte confirmar o valor da diversificação, a TMP suscita mais perguntas do que as responde. Será que a variância é mesmo uma medida boa de “risco”?³ Afinal, ela inclui oscilações tanto acima quanto abaixo do valor médio, e os investidores raramente se preocupam com o primeiro. Será que a TMP não pode nos dar uma medida melhor do risco – como a chance de o portfólio perder alguma porcentagem do seu valor? Em teoria, pode, se assumirmos que os retornos seguem alguma distribuição de probabilidade. Mas qual delas, e como podemos saber quando não funciona mais?

Aí há o problema dos valores dos retornos, da variância e das correlações com que alimentamos as equações para todos os ativos. Se estivéssemos lidando com a física, bastaria procurá-los em tabelas, e eles seriam constantes, como as massas de elétrons e prótons. Mas a única constante dos papéis financeiros é a constante mudança no seu retorno e volatilidade. É possível calcular valores médios – mas em que escala de tempo devem ser calculados, e o que acontece se seguirem distribuições nas quais a própria noção de variância não faz sentido?

Correlações são outra enorme fonte de incerteza; mesmo as regras práticas, como o fato de os papéis não se correlacionarem às ações, valem até o dia em que não valem mais. Pior: a mentalidade de rebanho dos investidores significa que papéis anticorrelacionados muitas vezes entram em sincronia justamente quando sua diversidade é mais necessária, como durante uma crise financeira.⁴

Diante de todos esses desafios, muitos investidores têm achado difícil confiar na matemática da TMP – entre eles o próprio Markowitz. Pouco tempo depois de desenvolver a teoria, ele foi confrontado com a necessidade de montar sua própria carteira de aposentadoria. Deveria ter analisado os registros de performance e calculado o mix ideal, mas descobriu que não podia encarar a

perspectiva de estar errado – e simplesmente pôs metade do seu dinheiro em ações e a outra metade em letras de renda fixa.⁵

Nas décadas após o seu surgimento, houve muitas tentativas de tornar a TMP mais sofisticada. O resultado tem sido uma enorme bibliografia técnica, mas pouca melhora além da ideia central de que a diversificação faz sentido. No final, nenhum volume de matemática pode dar à TMP – nem a qualquer estratégia de investimento – a confiabilidade da física. Elas sempre permanecerão modelos de fenômenos incertos cuja validade por si só é incerta. Em anos recentes, isso tem dado peso ao argumento de que simplesmente não há sentido em tentar criar portfólios e aplicar “gerência ativa”, comprando, vendendo e manejando o mix numa tentativa de se sair melhor que o mercado de ações como um todo. Essa é uma crença respaldada pela evidência de que muitos investidores “ativos” aparentemente bem-sucedidos são – como os fundos de cobertura – nada mais que pontos atípicos da curva cuja performance regride à média.⁶ Mesmo aqueles que conseguem bater o mercado fracassam em fazê-lo, com uma margem capaz de justificar as taxas de administração cobradas.⁷

Tudo isso tem levado alguns dos cérebros mais sagazes em finanças a argumentar que a melhor estratégia é a mais simples. David Swensen, superintendente do fundo de dotação de US\$ 24 bilhões da Universidade Yale, o proeminente analista quantitativo Paul Wilmott e a lenda dos investimentos, Warren Buffett, todos expressaram entusiasmo com portfólios que simplesmente imitam a performance do mercado usando os chamados fundos de “índice rastreador”.⁸ Como o nome sugere, eles são montados para rastrear o fluxo e refluxo dos índices do mercado como US S&P500, UK FTSE 100 ou MSCI World Allcap usando computadores. Esses fundos “passivos” nunca podem ter desempenho melhor que o seu índice, porém isso precisa ser ponderado com o fato de que o S&P500 tem apresentado a respeitável média de 8% ao ano de crescimento real desde 1985 – e os fundos ativos em geral falham para conseguir o mesmo desempenho, a despeito das taxas que lhes permitem tentar. Tampouco os fundos passivos nos livram da tarefa de diversificar; de modo geral, são necessários inúmeros deles, cobrindo diferentes áreas, para fazer frente à pior

volatilidade. Mas, com pouca intervenção humana, eles cobram taxas de administração muito baixas – a pior ameaça à performance de um portfólio.

A abordagem passiva também trata indiscutivelmente da fonte mais importante do mau desempenho dos investimentos: nós mesmos. Muita gente considera investir apenas uma forma de jogo mais elevada. Despejar dinheiro num punhado de papéis com diversificação zero decerto justificaria essa visão. Mas o jogo também é conhecido pela maneira como afeta a mente, com sua natureza probabilística, deflagrando uma legião de comportamentos potencialmente desastrosos: correr risco demais quando ganhamos; correr atrás de lucros quando não ganhamos; persistir com estratégias erradas sem nenhuma tentativa de avaliar o sucesso ou o fracasso. As incertezas inerentes de investir são conhecidas por afetar a mente de maneira semelhante. Estudos sugerem que a maioria dos investidores responde à natureza probabilística se tornando confiante demais ou complacente demais com sua própria atitude.⁹ Isso leva a uma miríade de comportamentos destrutivos de riqueza: atribuir alguns sucessos resultantes da sorte a uma habilidade genuína; apostar em “ganhos” onerosos que depois regredem à média; tomar erroneamente “ruído” de curto prazo por conhecimento de longo prazo. Investidores bem-sucedidos – como os jogadores profissionais – encontraram meios de controlar esses comportamentos e devem estar a quilômetros de distância da maioria de nós, especialmente nos tempos de crise. Nesse caso o melhor é fazer o mínimo possível. A maneira consagrada de realizar isso é via investimento de “comprar e segurar”, o que significa decidir o nosso portfólio e deixá-lo em paz.

Há evidência de sobra de que muitos podem fazer – e fazem – pior.¹⁰ Um estudo recente dos registros dos investidores de fundos mútuos nos Estados Unidos revelou que sua tendência de comprar ações que estão estourando e vender ações que estão desmoronando lhes custa muito caro. Entre 2000 e 2012, aqueles que tentaram identificar ações vencedoras e perdedoras tiveram um ganho médio anual de 3,6%. Em contraste, aqueles que simplesmente compraram e seguraram seu portfólio ganharam 5,6%.¹¹ Dois por cento ao ano pode não parecer muito, mas, mantendo-se ao longo de algumas décadas,

formariam uma taxa composta de 77% de aumento no valor do portfólio. Talvez a evidência mais convincente da abordagem de “fazer menos é mais” vem do mais celebrado dos investidores, Warren Buffett. Em uma de suas famosas cartas aos acionistas, ele revelou uma das pedras angulares de seu sucesso em lidar com os riscos e incertezas de investir: “Letargia beirando a preguiça.”¹²

Então, comprar e manter rastreadores de índices é o caminho para os investimentos bem-sucedidos? Decerto há evidência de que isso *pode* funcionar, mas no final é somente um *modelo* para investimentos de sucesso, e significa que sua abordagem para atacar a incerteza está ela própria sujeita à incerteza. Por exemplo, nos quinze anos desde o começo de 1985 até o fim de 1999, o S&P500 teve um impressionante índice de crescimento anual médio de 15% em termos reais. Investidores passivos se deram bem como bandoleiros, vendo o valor de seu portfólio aumentar oito vezes. No entanto, aqueles que assinaram a agenda passiva passaram os últimos quinze anos lutando para conseguir um crescimento médio de 2% ao ano, e acabaram com os portfólios apenas 30% mais valiosos. O que fizeram de errado? Nada – exceto falhar em prever que o modelo passivo estava prestes a decepcioná-los com os dois piores colapsos dos últimos cem anos: o estouro da bolha da informática (Dotcom Bubble) em 2000 e a crise financeira de 2007-08. Durante essas duas épocas, o modelo passivo exigiu que os que nele acreditavam simplesmente ficassem sentados assistindo ao colapso de seus portfólios. Enquanto isso, muitos veteranos do modelo ativo foram capazes de recorrer à sua experiência, preservar valor à medida que os índices despencavam, identificar pechinchas e dar a volta por cima.

ISSO FECHA O CÍRCULO da nossa análise da chance e da incerteza, e de como lidar com sua miríade de manifestações. A única regra importantíssima é esta: nunca perder de vista o fato de que, se, por habilidade ou sorte, acharmos a “coisa certa”, sempre existe uma chance de que ela nos decepcione. Nossa relutância em aceitar isso já provocou intermináveis sofrimentos, recriminação e culpa. Todavia, devemos nos autoflagelar apenas se falharmos em considerar as alternativas se a “coisa certa” der errado. Tudo o que podemos fazer sempre é

dar a nós mesmos a melhor *chance* de sucesso, aceitar que ela é sempre menor que 100% e nos preparar para essa eventualidade.

No final, devemos todos jogar os dados e correr os riscos.

^m Geeks: o termo já está consagrado em português, referindo-se aos “viciados” em computadores e em tudo que diz respeito ao mundo virtual. (N.T.)

Notas

1. O lançador de moedas prisioneiro dos nazistas

1. J.E. Kerrich, *An Experimental Introduction to the Theory of Probability*, Copenhagen, E. Munkgaard, 1946.
2. J. Strzałko et al., “Dynamics of coin tossing is predictable”, *Physics Reports*, v.469, n.2, 2008, p.59-92.
3. P. Diaconis et al., “Dynamical bias in the coin toss”, *Siam Review*, v.49, n.2, 2007, p.211-35.

3. O obscuro segredo do teorema áureo

1. S. Stigler, “Soft questions, hard answers: Jacob Bernoulli’s probability in historical context”, *Intl. Stat. Rev.*, v.82, n.1, 2014, p.1-16.
2. Aqui uma analogia pode ajudar. Arqueiros de primeira linha confiam que vão chegar perto do centro do alvo com apenas algumas setas. Em contraste, os principiantes têm baixa confiança de chegar perto do centro do alvo com o mesmo número de setas. Dando-lhe tempo suficiente, porém, mesmo eles podem ter muita confiança de que irão acertar algumas setas perto do centro do alvo. A questão sobre a qual Bernoulli lançou luz era: qual a relação entre o nível de confiança, a proximidade do centro do alvo e o número de tentativas?
3. Stigler, “Soft questions, hard answers”, op.cit.
4. Bernoulli havia tentado simplificar os cálculos ao usar seu teorema, mas eles eram crus demais. De Moivre encontrou aproximações melhores, e no processo inventou uma primeira versão do teorema do limite central, que iremos ver adiante.

5. Quais são as chances *disso*?

1. Para qualquer atributo (por exemplo, aniversário ou signo astrológico) em que todo mundo tenha chance igual de estar em um dos grupos G ($G = 365$ para aniversários, 12 para signos astrológicos), é necessária uma multidão de N pessoas para haver igual chance de pelo menos uma coincidência exata, onde N é 1, dezoito vezes a raiz quadrada de G . Para a teoria de outras coincidências, ver R. Matthews e F. Stones, “Coincidences: the truth is out there”, *Teaching Statistics*, v.20, n.1, 1998, p.17-9.

6. Pensar de modo independente não inclui gema de ovo

1. M. Hanlon, “Eggs-actly what ARE the chances of a double-yolker?”, *Daily Mail Online*, 3 fev 2010.

8. Aviso: há muito X por aí

1. J.A. Finegold et al., “What proportion of symptomatic side-effect in patients taking statins are genuinely caused by the drug?”, *Euro. J. Prev. Cardiol.*, v.21, n.4, 2014, p.464-74.
2. R. Matthews, “Medical progress depends on animal models – doesn’t it?”, *J. Roy. Soc. Med.*, v.101, n.2, 2008, p.95-8.

9. Por que o espetacular tantas vezes vira “mais ou menos”

1. B.G. Malkiel, resultados do estudo *Vanguard* citados em B.I. Murstein, “Regression to the mean: one of the most neglected but important concepts in the Stock Market”, *J. Behav. Fin.*, v.4, n.4, 2003, p.234-7.

10. Se você não sabe, vá pelo aleatório

1. D.A. Graham, “Rumsfeld’s knowns and unknowns: the intellectual history of a quip”, *The Atlantic* (online), 27 mar 2013.
2. R.A. Fisher, *the Design of Experiments*, Edimburgo, Oliver & Boyd, 1935, p.44.
3. I. Chalmers, “Why the 1948 MRC trial of streptomycin used treatment allocation based on random numbers”, *JLL Bulletin: “Commentaries on the history of treatment evaluation”*, 2010.
4. B. Djulbegovic et al., “Treatment success in cancer”, *Arch. Int. Med.*, n.168, 2008, p.632-42.
5. J. Henrich, S.J. Heine e A. Norenzayan, “The weirdest people in the world?”, *Behav. & Brain Sci.*, v.33, n.2, 2010, p.61-83.
6. P.M. Rothwell, “Factors that can affect the external validity of randomized controlled trials”, *PLOS Clin. Trials*, v.1, n.1, 2006, p.e9.
7. U. Dirnagl e M. Lauritzen, “Fighting publication bias”, *J. Cereb. Blood Flow & Metab.*, n.30, 2010, p.1263-4.
8. C.W. Jones e T.F. Platts-Mills, “Understanding commonly encountered limitations in clinical research: an emergency medicine resident’s perspective”, *Annals Emerg. Med.*, v.59, n.5, 2012, p.425-31.
9. S. Parker, “The Oportunidades Program in Mexico”, Shanghai Poverty Conference, 2003.
10. A. Petrosino et al., “ ‘Scared Straight’ and other juvenile awareness programs for preventing juvenile delinquency”, *Cochrane Database of Systematic Reviews*, n.4, 2013.
11. Por exemplo, o Behavioural Insights Team trabalha com o Cabinet Office do Reino Unido em abordagens da “teoria da cutucada” para implementação de políticas. Muitos de seus sucessos devem-se ao extensivo uso de ECRs; disponível em: tinyurl.com/Organ-Donation-Strategy.

11. Nem sempre é ético fazer a coisa certa

1. O site Behind the Headlines, do Serviço Nacional de Saúde do Reino Unido, presta um grande serviço descreditando essas alegações; ver, por exemplo: tinyurl.com/SleepingPillsAlzheimers.
2. World Cancer Research Fund International, “Diet, nutrition, physical activity and liver cancer”, relatório do Continuous Update Project, 2015.
3. J.N. Hirschhorn et al., “A comprehensive review of genetic association studies”, *Genetics in Medicine*, v.4, n.2, 2002, p.45-61.
4. R. Sinha et al., “Meat intake and mortality: a prospective study of over half a million people”, *Arch. Int. Med.*, v.169, n.6, 2009, p.562-71; M. Nagao et al., “Meat consumption in relation to mortality from cardiovascular disease among Japanese men and women”, *Euro. J. Clin. Nutr.*, v.66, n.6, p.687-93; S. Rohrmann et al., “Meat consumption and mortality-results from the European Prospective Investigation into Cancer and Nutrition”, *BMC Med.*, v.11, n.1, 2013, p.63.
5. S.S. Young e A. Karr, “Deming, data and observational studies: a process out of control and needing fixing”, *Significance*, set 2011, p.122-6.
6. M. Belson, B. Kingsley e A. Holmes, “Risk factors for acute leukemia in children: a review”, *Env. Health Persp.*, 2007, p.138-45.
7. A.B. Hill, “The environment and disease: association or causation?”, *Proc. Roy. Soc. Med.*, v.58, n.5, 1965, p.295-300.

12. Como uma “boi-bagem” deflagrou uma revolução

1. K. de Bakker, A. Boonstra e H. Wortmann, “Does risk management contribute to IT project success? A meta-analysis of empirical evidence”, *Intl. J. Proj. Mngt*, n.28, 2010, p.493-503; D. Ramel, “New analyst reports rips Agile”, *ADT Magazine*, 13 jul 2012; R. Bacon e C. Hope, *Conundrum: Why Every Government Gets Things Wrong and what We do About It*, Londres, Biteback, 2013.
2. Um dos mais conhecidos é o efeito Bradley, batizado com o nome do candidato democrata homônimo, indicado na eleição de 1982 para governador da Califórnia. Desde então afirma-se que ele desempenha seu papel em fracassos de pesquisas de opinião tais como as eleições gerais no Reino Unido de 1992 e 2015. Ironicamente, o efeito Bradley ocorreu provavelmente por um erro simples de amostragem: sua derrota foi dentro de 1%, facilmente atribuível aos “Não sei” – outra fonte de erro em pesquisas de opinião convencionais.
3. L. Hong e S.E. Page, “Groups of diverse problem solvers can outperform groups of high-ability problem solvers”, *PNAS*, n.101, 2004, p.16385-9.
4. C.P. Davis-Stober et al., “When is a crowd wise?”, *Decision*, v.1, n.2, 2014, p.79-101.
5. A.B. Kao e I.D. Couzin, “Decision accuracy in complex environments is often maximized by small group sizes”, *Proc. Roy. Soc. B*, v.281, n.1784, 2014, p.20133305.
6. S.M. Herzog e R. Hertwig, “Think twice and then: combining or choosing in dialectical bootstrapping?”, *J. Exp. Psychol.: Learning, Memory, and Cognition*, v.40, n.1, 2014, p.218-33.

14. Onde os espertinhos se dão mal

1. Mais sobre a teoria dos jogos de cassino e uma legião de outros aspectos da probabilidade pode ser encontrado no meu texto favorito sobre o assunto: John Haigh, *Taking Chances*, Oxford, Oxford University Press, 2003.

15. A regra áurea das apostas

1. J. Rosecrance, “Adapting to failure: the case of horse race gamblers”, *J. Gambling Behav.*, v.2, n.2, 1986, p.81-94.
2. P. Veitch, *Enemy Number One*, Newbury, Racing Post Books, 2010.

17. Fazer apostas melhores no cassino da vida

1. Seja P a chance de o boato se revelar verdadeiro, então a chance de o boato se revelar falso é $1 - P$ (como um desses dois resultados *deve* ser verdadeiro, suas chances devem somar 1). As consequências esperadas de ficar no lugar são então $-10P + 7(1 - P)$, enquanto a de se mudar são $2P + (1 - P)$. Igualando as duas expressões, obtemos a probabilidade acima da qual se mudar produz consequências mais positivas. Descobrimos então que faz sentido se mudar se a chance P de os boatos serem verdade exceder $\frac{1}{3}$.

18. Diga a verdade, doutor, quais as minhas chances?

1. Alice Thomson é o pseudônimo de uma pessoa real, contatada pelo autor em janeiro de 2015.
2. G. Gigerenzer, in *Reckoning with Risk*, Londres, Allen Lane, 2002, p.42-5.
3. K. Moisse, “Man takes pregnancy test as joke, finds testicular tumor”, *ABC News online*, 6 nov 2012.
4. Essa é uma simples consequência do teorema de Bayes, descrito no Capítulo 20.

19. Isso não é uma simulação! Repito, isso não é uma simulação!

1. R. Matthews, “Decision-theoretic limits on earthquake prediction”, *Geophys. J. Int.*, n.131, n.3, 1997, p.526-9.
2. Idem, “Base-rate errors and rain forecasts”, *Nature*, n.382, 1996, p.766.

20. A fórmula milagrosa do reverendo Bayes

1. Baseado in Paul Tough, “A speck in the sea”, *New York Times Magazine*, 2 jan 2014.
2. Para um relato facilitado acerca do teorema de Bayes, sua história e suas aplicações, ver S.B. McGrayne, *The Theory That Would Not Die*, New Haven, Yale University Press, 2011.
3. Disponível em: tinyurl.com/Bayes-Essay.

4. As fórmulas vêm da chamada distribuição binomial.
5. Ao longo do livro, mantenho o foco na forma mais simples do teorema, envolvendo uma dicotomia direta entre uma hipótese e todas as alternativas. Deve-se salientar, porém, que o teorema de Bayes pode lidar com casos bem mais complexos.
6. Para uma análise cuidadosamente debatida da luta de Bayes com o “problema dos a priori” e as concepções errôneas que se seguiram, ver S.M. Stigler, “Thomas Bayes’s bayesian inference”, *Journal of the Royal Statistical Society, Series A (General)*, 1982, p.250-8.
7. Contrariamente ao que até muitos defensores do raciocínio bayesiano pensam, porém, a mesma evidência pode separar muito mais os dois campos numa controvérsia. Ver R. Matthews, “Why do people believe weird things?”, *Significance*, dez 2005, p.182-4.

21. O encontro do dr. Turing com o reverendo Bayes

1. I.J. Good, “Studies in the history of probability and statistics. XXXVII: AM Turing’s statistical work in World War II”, *Biometrika*, 1979, p.393-6.
2. S. Zabbell, “Commentary on Alan M. Turing: the applications of probability to cryptography”, *Cryptologia*, v.36, n.3, 2012, p.191-214.
3. Y. Suhov e M. Kelbert, *Probability and Statistics by Example, v.2: Markov Chains: A Primer in Random Processes and Their Applications*, Cambridge, Cambridge University Press, 2008, p.433.
4. Essa é consequência da aplicação de logaritmos ao original. A fórmula resultante não aparece explicitamente no relatório de Turing, mas “transformação logarítmica” é uma parte fundamental de seus argumentos.
5. D.A. Berry, “Bayesian clinical trials”, *Nat. Rev. Drug. Discov.*, v.5, n.1, 2006, p.27-36.
6. M. Dembo et al., “Bayesian analysis of a morphological supermatrix sheds light on controversial hominin relationships”, *Proc. R. Soc. B.*, v.282, n.1812, 2015, 20150943.
7. R. Trotta, “Bayes in the sky: bayesian inference and modal selection in cosmology”, *Contemp. Physics*, v.49, n.2, 2008, p.71-104.

22. Usando Bayes para julgar melhor

1. R. Matthews, “The interrogator’s fallacy”, *Bull. Inst. Math. Apps.*, v.31, n.1, 1995, p.3-5.
2. S. Connor, “The science that changed a minister’s mind”, *New Scientist*, 29 jan 1987, p.24.

23. Um escândalo de significância

1. H. Jeffries, *Theory of Probability*, 1939, p. 388-9; W. Edwards, H. Lindman e L.J. Savage, “Bayesian statistical inference for psychological research”, *Psychol. Rev.*, v.70, n.3, 1963, p.193-242; J. Berger e T. Sellke, “Testing a point null hypothesis: the irreconcilability of P-values and evidence”, *Jasa*, v.82,

- n.397, 1987, p.112-22; R. Matthews, “Why should clinicians care about Bayesian methods?”, *J. Stat. Plan. Infer.*, v.94, n.1, 2001, p.43-58; “Flukes and flaws”, *Prospect*, nov 1998.
2. Ver P.R. Band, N.D. Le, R. Fang e M. Deschamps, “Carcinogenic and endocrine disrupting effects of cigarette smoke and risk of breast cancer”, *Lancet*, v.360, n.9339, 2002, p.1044-9, contraditado no mês seguinte pelo Collaborative Group on Hormonal Factors in Breast Cancer, “Alcohol, tobacco and breast cancer”, *B.J. Canc.*, v.87, n.11, 2002, p.1234-45.
 3. Para uma interessante demonstração, ver G.D. Smith e E. Shah, “Data dredging, bias, or confounding: they can all get you into the *BMJ* and the Friday papers”, *BMJ*, v.325, n.7378, 2002, p.1437.
 4. G. Taubes, “Epidemiology faces its limits”, *Science*, v.269, n.5221, 1995, p.164-9.
 5. J.P.A. Ioannidis, “Why most published research findings are false”, *PLOS Medicine*, v.2, n.8, 2005, p.e124.
 6. Idem, “Contradicted and initially stronger effects in highly cited clinical research”, *Jama*, v.294, n.2, 2005, p.218-28; R.A. Klein et al., “Investigating variation in replicability: a ‘many labs’ replication project”, *Social Psychology*, v.45, n.3, 2014, p.142-52; M. Baker, “First results from psychology’s largest reproducibility test”, *Nature* online news, 30 abr 2015.
 7. 2014 Global R&D Funding Forecast (Bastelle.org, dez 2013).
 8. R.A. Purdy e S. Kirby, “Headaches and brain tumors”, *Neurol. Clin.*, v.22, n.1, 2004, p.39-53.
 9. J. Aldrich, “R.A. Fisher on Bayes and Bayes’ Theorem”, *Bayesian Analysis*, v.3, n.1, 2008, p.161-70.
 10. R.A. Fisher, “The statistical method in physical research”, *Proc. Soc. Psych. Res.*, n.39, 1929, p.189-92; Fisher descreve explicitamente a natureza arbitrária do valor padrão p para significância, e adverte sobre os perigos da má interpretação.
 11. F. Yates, “The influence of statistical methods for research workers on the development of the science of statistics”, *Jasa*, v.46, n.253, 1951, p.19-34.
 12. F. Fidler et al., “Editors can lead researchers to confidence intervals, but can’t make them think statistical reform lessons from medicine”, *Psych. Sci.*, v.15, n.2, 2004, p.119-26.
 13. S.T. Ziliak e D.N. McCloskey, *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice and Lives*, Ann Arbor, University of Michigan Press, 2008, cap.7.
 14. F.L. Schmidt e J.E. Hunter, “Eight common but false objections to the discontinuation of significance testing in the analysis of research data”, in L.L. Harlow et al. (orgs.), *What If There Were No Significance Tests?*, Oxford, Psychology Press, 1997, p.37-64.
 15. Quando o autor começou a fazer reportagens sobre esses temas, na década de 1990, foi-lhe dito por diversos organismos estudados, inclusive a Royal Statistical Society e a British Psychological Society, que afirmações claras acerca de política sobre valores p eram desafiadoras demais para seus membros e publicações científicas.
 16. D.M. Windish. S.J. Huot e M.L. Green, “Medicine residents’ understanding of the biostatistics and results in the medical literature”, *Jama*, v.298, n.9, 2007, p.1010-22.

24. Esquivando-se da espantosa máquina de bobagens

1. J. Maddox, “Cern comes out again on top”, *Nature*, v.310, n.97, 12 jul 1984.
2. J.W. Moffat, *Cracking the Particle Code of the Universe*, Oxford, Oxford University Press, 2014, p.113.
3. Valores sigma são uma medida do grau de separação entre os resultados obtidos e o que seria de esperar se não fossem nada além de casualidades. Logo, diferentemente dos valores p , quanto *maior* o valor sigma, maior a separação entre os resultados e as meras casualidades. São também medidas altamente não lineares de “significância”, em que um salto de sigma de 2 para 4 corresponde a um aumento de 700 vezes na “significância”. Voltaremos a encontrá-los ao tratarmos da crise financeira.
4. D. Mackenzie, “Vital statistics”, *New Scientist*, 26 jun 2004, p.36-41.
5. Ver, por exemplo, R. Matthews, “Why should clinicians care about Bayesian methods?”, *JSPI*, v.94, n.1, 2001, p.43-58.
6. Os resultados citados baseiam-se, na teoria, in J. Berger e T. Sellke, “Testing a point null hypothesis: the irreconcilability of P-values and evidence”, *Jasa*, v.82, n.397, 1987, p.112-22 (especialmente seção 3.5); dadas as premissas sobre distribuição e os limites inferiores envolvidos, as cifras são somente indicativas.

25. Use aquilo que você já sabe

1. W.W. Rozeboom, “Good Science is abductive, not hypothetico-deductive”, in L.L. Harlow et al. (orgs.), *What If There Were No Significance Tests?*, Oxford, Psychology Press, 1997, p.335-92.
2. Em termos simples, o problema reside no fato de que muitas questões de pesquisa envolvem intervalos (“distribuições”) de probabilidades a priori e também de explicações alternativas dos dados. Em casos simples, podem-se usar “densidades conjugadas”, dando fórmulas nas quais se inserem dados e conhecimentos a priori, mas muitas aplicações da vida real demandam técnicas intensivas de computação.
3. S. Connor, “Glaxo chief: our drugs do not work on most patients”, *Independent*, 8 dez 2003, p.1.
4. S.J. Pocock e D.J. Spiegelhalter, “Domiciliary thrombolysis by general practitioners”, *BMJ*, v.305, n.6860, 1992, p.1015.
5. Por si só, o IC de 95% significa que, se pegássemos uma grande amostra aleatória (nesse caso, de participantes do experimento) retirada da mesma população (nesse caso, todos os pacientes apropriados), poderíamos ter confiança de que o IC resultante cobriria o valor da população do que quer que nos interessasse – digamos, uma taxa de risco de morte – 95% das vezes (presumindo, claro, que todas as fontes de erro não aleatórias, como um viés, tenham sido eliminadas). Logo, a “confiança” está relacionada à confiabilidade da *técnica estatística*, e não à confiabilidade do *achado*. Bayes mostra que podemos pegar a primeira como medida da segunda apenas se estivermos em estado de absoluta ignorância daquilo que o achado poderia ser – o que raramente acontece. Depois de décadas de pesquisa, geralmente temos conhecimento a priori ao qual recorrer, e Bayes então nos dá um intervalo *crível* de 95%, onde a credibilidade realmente se relaciona com o achado (com as condições habituais de que o experimento esteja livre de outras fontes de erro).

6. L.J. Morrison et al., “Mortality and pre-hospital thrombolysis for acute myocardial infarction: a meta-analysis”, *Jama*, v.283, n.20, 2000, p.2686-92.

26. Desculpe, professor, mas não engulo essa

1. S. Kühn e J. Gallinat, “Brain structure and functional connectivity associated with pornography consumption”, *Jama Psychiatry*, v.71, n.7, 2014, p.827-34.
2. J.A. Tabak e V. Zayas, “The roles of featural and configural face processing in snap judgements of sexual orientation”, *Plus One*, v.7, n.5, 2012, e36671.
3. I. Chalmers e R. Matthews, “What are the implications of optimism bias in clinical research?”, *The Lancet*, v.367, n.9509, 2006, p.449-50. Para os desafios de “elicitación a priori” em experimentos clínicos, ver D.J. Spiegelhalter, K.R. Abrams e J.P. Myles, *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, Chichester, Wiley, 2004, p.147-8. Seres humanos em geral parecem ser enviesados no sentido de uma visão rósea de eventos futuros; ver, por exemplo, T. Sharot, “The optimism bias”, *Current Biology*, v.21, n.23, 2011, p.R941-5.
4. R. Matthews, “Methods for assessing the credibility of clinical trial outcomes”, *Drug Inf. Ass. J.*, v.35, n.4, 2001, p.1469-78; disponível em: tinyurl.com/credibility-prior; aqui há disponível uma calculadora on-line: statpages.org/bayecred.html.
5. H. Gardener et al., “Diet soft drink consumption is associated with an increased risk of vascular events in the Northern Manhattan Study”, *J. Gen. Int. Med.*, v.27, n.9, 2012, p.1120-6.
6. Os trabalhos de Ramsey e De Finetti nos anos 1920 e de Cox e Jaynes nos anos 1960 apontavam para a inelutabilidade do cálculo de probabilidades para captar uma crença; ver C. Howson e P. Urbach, *Scientific Reasoning: The Bayesian Approach*, Chicago, Open Court, 1993, cap. 5.
7. K.H. Knuth e J. Skilling, “Foundations of inference”, *Axioms*, v.1, n.1, 2012, p.38-73.

27. A assombrosa curva para tudo

1. R.J. Gillings, “The so-called Euler-Diderot incident”, *Am. Math. Monthly*, v.61, n.2, 1954, p.77-80.
2. E. O’Boyle e H. Aguinis, “The best and the rest: revisiting the norm of normality of individual performance”, *Personnel Psych*, n.65, 2012, p.79-119; J. Bersin, “The myth of the Bell curve: look for the hyper-performers”, *Forbes online*, 19 fev 2014.
3. Para eventos cuja probabilidade num experimento único é p ($= 0,5$ para o lançamento de uma moeda), as chances de obter S sucessos em qualquer ordem durante x tentativas são dadas pela distribuição binomial: $[S!/(S - x)!x!]p^x(1 - p)^{S - x}$, onde $!$ significa fatorial, que pode ser encontrado em qualquer calculadora científica. Assim, as chances de obter exatamente cinco caras em dez lançamentos é $[10!5!5!](0,5)^5(1 - 0,5)^{10 - 5} = 0,246$. Os fatoriais e as potências tornam-se muito tediosos para se trabalhar num S grande.
4. Estritamente falando, a versão “clássica” de Laplace do teorema também impõe restrições ao comportamento dessas influências aleatórias independentes. Os matemáticos, desde então, provaram que

o teorema continua valendo – e a curva do sino acaba surgindo – mesmo quando as influências aleatórias não se comportam de maneira idêntica. Todavia, mesmo sob as chamadas condições de Lindenberg-Feller, as influências devem ser independentes e incapazes de assumir um comportamento louco demais – o que constitui ainda uma limitação importante.

5. Tais argumentos são frequentemente questionáveis: ver A. Lyon, “Why are normal distributions normal?”, *B. J. Phil. Sci.*, v.65, n.3, 2014, p.621-49.
6. S. Stigler, *Statistics on the Table*, Cambridge, MA, Harvard University Press, 2002, p.53.
7. *Ibid.*, p.412.
8. In H. Jeffreys, “The law of error in the Greenwich variation of latitude observations”, *Mon. Not. RAS*, v.99, n.9, 1939, p.703.

28. Os perigos de pensar que tudo é normal

1. K. Dowd et al., “How unlucky is 25-sigma?”, pré-impressão in ArXiv.org: arXiv:1103.5672, 2011.
2. K. Pearson, “Notes on the history of correlation”, *Biometrika*, n.13, 1920, p.25-45.
3. Isso baseia-se em dados da vida real do Censo dos Estados Unidos de 1999, relatado e analisado por M.F. Schilling, A.E. Watkins e W. Watkins, “Is human height bimodal?”, *Am. Stat.*, v.56, n.3, 2002, p.223-9.
4. Como mostram Schilling et al. (*ibid.*), se a diferença entre as médias de diversas curvas do sino exceder um certo múltiplo da soma dos desvios-padrão, a curva do sino combinada terá um aspecto distintamente dentado. Adicionar mais curvas do sino tende também a distorcer o formato inteiro da curva composta, estragando sua simetria.
5. Ver, por exemplo, R.W. Fogel et al., “Secular changes in American and British stature and nutrition”, *J. Interdis. Hist.*, v.14, n.2, 1983, p.445-81.
6. Ver, por exemplo, B. Mandelbrot, *The Misbehaviour of Markets*, Londres, Profile, 2005, que se mostrou presciente no acompanhamento da crise financeira. Para um relato das consequências, ver A.G. Haldane e B. Nelson, “Tails of the unexpected”, nas atas de *The Credit Crisis Five Years On: Unpacking the Crisis*, Escola de Negócios da Universidade de Edimburgo, 8-9 jun 2012.
7. P. Wilmott, “The use, misuse and abuse of mathematics in finance”, *Phil. Trans. Roy. Soc.*, Série A, v.358, n.1765, 2000, p.63-73.
8. Incluem o chamado método do valor em risco (VAR), desenvolvido por engenheiros das finanças no fim dos anos 1980, e agora parte dos chamados padrões internacionais de risco bancário, Basileia III, produzidos após a crise financeira. O VAR envolve a estimativa por parte das instituições financeiras das chances de ter perdas específicas num contexto de tempo específico. Tais estimativas baseiam-se muitas vezes em dados históricos e simulações, que carregam riscos óbvios. Foram sonoramente atacados por Nassim Taleb, autor de *The Black Swan* (Londres, Penguin, 2008); ver, por exemplo, www.fooledbyrandomness.com/jorion.html.
9. JPMorgan Chase, *Annual Report*, abr 2014, p.31.

29. Irmãs feias e gêmeas malvadas

1. Isso é verdade para qualquer distribuição simétrica – isto é, presumindo que a média exista –, o que, como veremos com a distribuição de Cauchy, pode não acontecer.
2. D. Veale et al., “Am I normal? A systematic review and construction of nomograms for flaccid and erect penis length and circumference in up to 15 521 men”, *BJU Intl.*, v.115, n.6, 2015, p.978-86.
3. O. Svenson, “Are we all less risky and more skillful than our fellow drivers?”, *Acta Psychol.*, v.47, n.2, 1981, p.143-8.
4. S. Powell, “RAC Foundation says young drivers more likely to crash”. *BBC Newsbeat*, 27 mai 2014.
5. Isso se reflete no uso de logaritmos. O teorema do limite central de Laplace mostra que obtemos uma curva normal-padrão como resultado de influências aleatórias independentes que se somam. O uso de logaritmos retém esta propriedade aditiva para influências que na realidade agem de forma multiplicativa.
6. E. Limpert, W.A. Stahel e M. Abbt, “Log-normal distributions across the sciences: keys and clues”, *BioScience*, v.51, n.5, 2001, p.341-52.
7. Ver L.T. DeCarlo, “On the meaning and use of kurtosis”, *Psych. Meth.*, v.2, n.3, 1997, p.292-307.
8. Alguns textos avançados ressaltam que se pode criar inadvertidamente uma distribuição de Cauchy formando a razão de duas variáveis com distribuição normal, onde o denominador passa pelo zero. Isso pode causar um estrago irreconhecível mesmo no cálculo de características básicas como a média e o desvio-padrão dessa razão, para não mencionar o “teste de significância”.
9. Usando a teoria da curva do sino, é possível mostrar que um evento 25-sigma tem uma probabilidade impressionantemente baixa de 1 em 10^{137} , ou seja, 1 seguido de 137 zeros. Segundo a distribuição de Cauchy, porém, as chances são de 1 em 77, em outras palavras, cerca de 10^{135} vezes mais provável que sugere o cálculo da curva do sino. Nunca se deve esquecer que eventos incrivelmente raros podem acontecer e acontecem o tempo todo; as chances de você ter outras 24 horas precisamente iguais àquelas que acabou de passar são muito menores que 10^{137} . Mas, então, nenhuma pessoa sã tenta desenvolver uma teoria capaz de prever essas coisas; em finanças, elas o fazem.
10. E.F. Fama, “The behavior of stock-market prices”, *J. Business*, v.38, n.1, 1965, p.34-105.
11. Batizadas em homenagem a Paul Lévy (1886-1971), às vezes também são chamadas distribuições estáveis paretianas – ou simplesmente “estáveis”. Fama veio a empregá-las depois de tomar conhecimento do trabalho de Benoit Mandelbrot.
12. Seu comportamento pode ser sintonizado utilizando quatro “botões de controle” – parâmetros, no jargão – que determinam a localização de pico, achatamento, distorção e – o mais importante – a “grossura” das caudas. Esta última é ditada por um número entre zero e 2. Quando é exatamente 2, o resultado é a curva do sino, porém, para valores mais baixos, as distribuições têm variância infinita. Quando chega exatamente a 1 torna-se a curva de Cauchy, carecendo tanto de média quanto de variância. Valores abaixo de 1 dão resultados insanos.
13. Para um relato não técnico, com muitos exemplos e insights da vida real, ver M.E.J. Newman, “Power laws, Pareto distributions and Zipf’s law”, *Contemp. Physics*, v.46, n.5, 2005, p.323-51.

14. A ameaça apresentada pelas leis de potência à confiabilidade da pesquisa na área de negócios é examinada em G.C. Crawford, W. McKelvey e B. Lichtenstein, “The empirical reality of entrepreneurship: how power law distributed outcomes call for new theory and method”, *J. Bus. Vent. Insight*, v.1, n.2, 2014, p.3-7.

30. Até o extremo

1. R.A. Fisher e L. Tippett, “Limiting forms of the frequency distribution of the largest or smallest member of a sample”, *Math. Proc. Camb. Phil. Soc.*, v.24, n.2, 1928, p.180-90.
2. Essas regras práticas emergem naturalmente das distribuições de lei de potência. Uma distribuição de lei de potência da forma $p(x) = Cx^{-a}$ leva a uma proporção X da quantidade total de uma grandeza (digamos, a riqueza do mundo) ligada a uma porcentagem P da população total, onde $X = P^K$ e $K = (a - 2)/(a - 1)$. Assim, por exemplo, $a = 2,2$ dá a famosa expressão de que “quase 80% da riqueza está concentrada nas mãos de apenas 20% da população mundial”.
3. M. Moscadelli, “The modelling of operational risk: experience with the analysis of the data collected by the Basel Committee”, *Temi di discussione* (Economic working papers), n.517, Bank of Italy Economic Research Department, 2004.
4. K. Aas, “The role of extreme value theory in modeling financial risk”, Conferência NTNU, Trondheim, 2008.
5. K. Aarssen e L. de Haan, “On the maximal life span of humans”, *Math. Pop. Studies*, v.4, n.4, 1994, p.259-81.
6. Em N experimentos de um evento aleatório de probabilidade P , o comprimento da maior sequência contínua é L , e satisfaz a equação $N(1 - P)^{P^L} = 1$. Ver M.F. Schilling, “The surprising predictability of long runs”, *Math. Mag.*, n.85, 2012, p.141-9.
7. M. Tsai e L. Chen, “The calculation of capital requirement using Extreme Value Theory”, *Economic Modelling*, v.28, n.1, 2011, p.390-5.

31. Assista a um filme de Nicolas Cage e morra

1. Ao contrário da crença disseminada, coeficientes de correlação não dizem nada sobre o tamanho da mudança produzida em uma variável por mudanças feitas na outra. Tampouco a correlação é apenas mensurável para relações simples, lineares: a correlação de Spearman é capaz de lidar com as relações não lineares monotônicas, e até com não normalidade.
2. Para conjuntos com pelo menos dez pares de dados, qualquer nível de correlação cuja magnitude absoluta exceda 0,62 será “estatisticamente significativo” melhor que o usual padrão $p = 0,05$. A maioria das correlações de Vigen, altamente anunciadas, passa esse padrão com facilidade – voltando a sublinhar as inadequações do conceito de “significância estatística” como meio de eliminar absurdos.
3. Se tudo isso já não fosse ruim o suficiente, o método mais usado para determinar correlações tem a premissa de comportamento de curva do sino embutido em sua própria essência.

4. A esquisita noção de cegonhas trazendo bebês aparece em “As cegonhas”, conto publicado em 1838 por Hans Christian Andersen, mas a mitologia parece bem mais antiga. Desde então ela foi “confirmada”, usando-se análise de correlação, por vários pesquisadores, incluindo o autor R. Matthews, “Storks deliver babies ($p = 0,008$)”, *Teaching Statistics*, v.22, n.2, 2000, p.36-8, que a utiliza para ilustrar as inadequações dos valores p ; ver também T. Höfer e H. Przyrembel, “New evidence for the theory of the stork”, *Paed. & Peri. Epid.*, v.18, n.1, 2004, p.88-92.
5. M.H. Meier et al., “Persistent cannabis users show neuropsychological decline from childhood to midlife”, *Pnas*, v.109, n.40, 2012, p. E2657-E2664.
6. O. Rogeberg, “Correlations between cannabis use and IQ change in the Dunedin cohort are consistent with confounding from socioeconomic status”, *Pnas*, v.110, n.11, 2013, p.4251-4.
7. Existe, por exemplo, evidência de que os riscos para a saúde do fumo passivo podem ser mais baixos do que frequentemente se alega; ver J.E. Enstrom, G.C. Kabat e G. Davey Smith, “Environmental tobacco smoke and tobacco related mortality in a prospective study of Californians, 1960-98”, *BMJ*, v.326, n.7398, 2003, p.1057-67. Este não é um tema acadêmico: se o risco deste fator de confusão comum for superestimado, pode fazer com que outras fontes de doenças respiratórias e cardíacas não sejam percebidas.
8. D. Freedman, R. Pisani e R. Purves, *Statistics*, 3ª ed., Nova York, W.W. Norton, 1998, p.149. O fenômeno da variância mutável faz a festa no termo *heterocedasticidade* (das palavras gregas para “diferente” e “dispersão”).
9. Sustentação para as preocupações de Pearson está in W. Dunlap, J. Dietz e J.M. Cortina, “The spurious correlation of ratios that have common variables: a Monte Carlo examination of Pearson’s formula”, *J. Gen. Psych.*, v.124, n.2, 1997, p.182-93. Para uma discussão do problema das correlações baseadas em proporções nos negócios, ver R.M. Wiseman, “On the use and misuse of ratios in strategic management research”, in D.D. Bergh e D.J. Ketchen (orgs.), *Research Methodology in Strategy and Management*, v.5, Bingley, Emerald Group Publishing, 2008, p.75-110.
10. Essas variações sazonais na temperatura são principalmente resultado da inclinação do eixo da Terra em relação à sua órbita em torno do Sol. Vale a pena ressaltar que *existem* técnicas para lidar com correlações não lineares, mas nem todo mundo que precisa delas sabe de sua existência – ou as emprega.

32. Temos de traçar a linha em algum lugar

1. Várias definições de “melhor” são possíveis, mas a regressão linear baseia-se no chamado princípio dos mínimos quadrados sugerido por Gauss, que possui algumas propriedades elegantes. A ideia básica é cometer o menor erro possível ao estimar uma variável usando outra.
2. J. Ginsberg et al., “Detecting influenza epidemics using search engine query data”, *Nature*, n.457, 2009, p.1012-4.
3. D. Lazer et al., “The parable of Google Flu: traps in big data analysis”, *Science*, n.343, 2014, p.1203-5.
4. C. Anderson, “The end of theory: the data deluge makes the scientific method obsolete”, *Wired*, 23 jun 2008.

5. O eminente estatístico britânico sir David Spiegelhalter, citado in T. Harford, “Big Data: are we making a big mistake?”, *Financial Times*, 28 mar 2014.
6. Levantamento de Gartner; disponível em: gartner.com/newsroom/id/2848718, 17 set 2014; valor de mercado tirado do relatório da *Forbes*, “6 predictions for the \$ 125 billion Big Data analytics Market in 2015”, publicado on-line, 11 dez 2014.
7. S. Finlay, *Predictive Analytics, Data Mining and Big Data*, Londres, Palgrave Macmillan, 2014, p.131.
8. Se os pares de dados (x, y) seguem uma relação de “lei de potência”, tal como $y = ax^n$, então $\log(y) = \log(a) + n\log(x)$, que é a fórmula para uma linha reta com intersecção vertical $\log(a)$ e inclinação n . A regressão linear aplicada aos pares de dados então fornece as “melhores” estimativas para $\log(a)$ e n – sendo este último a potência buscada.
9. P. Bak, *How Nature Works: The Science of Self-Organized Criticality*, Nova York, Springer, 1996.
10. Para uma análise abrangente tanto dos problemas teóricos quanto dos empíricos, ver A. Clauset, C.R. Shalizi e M.E.J. Newman, “Power-law distributions in empirical data”, *Siam Review*, v.51, n.4, 2009, p.661-703. Como na correlação, há maneiras de flexibilizar alguns dos “termos e condições” da regressão linear apresentada nos livros-texto, sobretudo os métodos “não paramétricos”, que funcionarão sem se conhecerem as distribuições subjacentes envolvidas. Mas estes ainda assim podem lutar com o comportamento selvagem das distribuições de lei de potência.
11. A.M. Edwards, “Overturning conclusions of Lévy flight movement patterns by fishing boats and foraging animals”, *Ecology*, v.92, n.6, 2011, p.1247-57.
12. N.E. Humphries et al., “Foraging success if biological Lévy flights recorded in situ”, *Pnas*, v.109, n.19, 2012, p.7169-74.

33. Jogar com os mercados não é uma ciência precisa

1. B. Keeley e P. Love, “Pensions and the crisis”, in *From Crisis to Recovery: The Causes, Course and Consequences of the Great Recession*, Paris, OECD Publishing, 2010.
2. Um contrato de derivativo para um mercador da Mesopotâmia é datado de 1809 a.C.; ver E.J. Weber, “A short history of derivative security markets”, Discussion Paper 08.10, Escola de Negócios da Universidade da Austrália Ocidental, 2008.
3. Exemplos eminentes incluem Emanuel Derman, Paul Wilmott e Riccardo Rebonato. Derman é um ex-físico de partículas da Universidade Columbia e autor de *Models. Behaving. Badly* (Nova York, Simon & Schuster, 2011). Wilmott é coautor, com Derman, de *Financial Modeller’s Manifesto*, e tem doutorado em dinâmica dos fluidos na Universidade de Oxford. Rebonato é autor do presciente *Plight of the Fortune Tellers* (Princeton University Press, 2007) e tem doutorado em física da matéria condensada.
4. A.W. Lo e M.T. Mueller, “Warning: physics envy may be hazardous to your wealth!”, *J. Invest. Mngt.*, v.8, n.2, 2010, p.13-63.
5. Como a resistência do ar varia com a velocidade do projétil, que por sua vez muda, em resposta, é necessário um cálculo avançado para determinar a trajetória. Acrescentem-se um alvo móvel e a rotação da Terra, e você tem a balística – o foco de pesquisa dos principais físicos na Segunda Guerra Mundial.

6. Um exemplo simples do uso da simetria é a descrição de um pedaço de papel quadrado; se o girarmos 90 graus, ele parece idêntico – “mudou sem ter mudado”. Simetrias mais sutis têm vínculos sutis com outros princípios poderosos da física: leis de conservação, sendo que o vínculo se manifesta por um espantoso resultado matemático conhecido como teorema de Noether.
7. Lo e Mueller, op.cit., seção 2.3.
8. Para uma análise de quais são essas habilidades e como podem ser implementadas nas finanças, ver ibid.

34. Cuidado com geeks criando modelos

1. A.W. Lo et al., “Hedge funds: a dynamics industry in transition”, *Ann. Rev. Fin. Econ.*, n.7, 2015.
2. Outro critério de medida frequentemente usado é a chamada *volatilidade* de um ativo, dada pela raiz quadrada da variância, conhecida em estatística como desvio-padrão.
3. Por exemplo, a correlação entre o índice de mercado US S&P500 e letras de longo prazo do Tesouro dos Estados Unidos trocou de sinal 29 vezes de 1927 a 2012, variando de -0,93 a +0,84. Ver N. Johnson et al., “The stock-bond correlation”, Pimco Quantitative Research Report, nov 2013.
4. Ver, por exemplo, N. Waki, “Diversification failed this year”, *New York Times Business*, 7 nov 2008; S. Stovall, “Diversification: a failure of fact or expectations?”, *Am. Ass. Indiv. Inv. J.*, mar 2010.
5. In J. Zweig, *Your Money and Your Brain: How the new science of neuroeconomics can help make you rich*, Nova York, Simon & Schuster, 2007, p.4.
6. R. Ferri, “Coin flipping outdoes active fund managers”, *Forbes*, 13 jan 2014.
7. Pesquisa do Departamento de Comunidades e Governo Local do Reino Unido, in M. Johnson, “We don’t need 80% of active management”, *Financial Times*, 11 mai 2014.
8. Disponível no blog Monevator, “The surprising investment experts who use index funds”, 10 fev 2015.
9. K.H. Baker e V. Ricciardi, “How biases affect investor behaviour”, *Euro. Fin. Rev.*, 28 fev 2014.
10. J. Kimelman, “The virtues of inactive investing”, *Barron’s*, 10 set 2014.
11. Y. Chien, “Chasing returns has a high cost for investors”, estudo do Federal Bank de St. Louis, 14 abr 2014.
12. Galas A., “Lethargy bordering on sloth: one of Warren Buffett’s best investing strategies”, *The Motley Fool*, 16 nov 2014.

Agradecimentos

A profundidade, amplitude e extensão das leis da probabilidade são assombrosas. Cada aspecto delas, desde sua história e interpretação até seus fundamentos teóricos e aplicações práticas, poderia formar a base de um livro de toda uma vida. De todas as disciplinas que enfrentei durante mais de trinta anos como estudioso e escritor na área científica, a probabilidade é aquela que continua a me intrigar e a me deixar determinado a aprender mais. Também descobri que ela tem o mesmo efeito sobre aqueles que a estudam e a usam profissionalmente – criando uma comunidade de pesquisadores e praticantes com uma inusitada mistura de características. Eles têm cérebro do tamanho de planetas, combinado a uma encantadora modéstia e disposição de ajudar qualquer um que tenha esperança de entender os caminhos da aleatoriedade, do risco e da incerteza. Foi um privilégio passar algum tempo na companhia deles ao longo dos anos, tirando proveito de sua experiência e do conhecimento. Quero agradecer especialmente a Doug Altman, Iain Chalmers, Steven Cowley, Peter Donnelly, Frank Duckworth, Gerd Gigerenzer, o saudoso Jack Good, John Haigh, Colin Howson, o saudoso Dennis Lindley, David Lowe, Paul Parsons, Peter Rothwell, Stephen Senn, David Spiegelhalter e Henk Tijms.

Este livro não existiria sem a sugestão inicial de Ian Stewart, o constante entusiasmo de John Davey, da Profile Books, o amor e apoio de Denise Best, minha companheira, musa e melhor amiga.

Quanto aos erros deste livro, todos são obra minha, e eu recebo de bom grado as correções. A experiência me ensinou que a probabilidade de eu cometer zero erro em questões de probabilidade é por si só zero.

Índice remissivo

administração de projetos, 1, 2-3

Affleck, Ben, 1

Agência Meteorológica do Reino Unido, 1, 2

agências de apostas, 1-2, 3-4, 5

aleatoriedade:

- Cern, trabalho da equipe do, 1

- definição, 1-2

- ignorância, 1-2, 3

- jogos de cassino, 1-2

- padrões, 1-2, 3-4

- pesquisa científica, 1-2

- pesquisa médica, 1-2, 3-4

- política de tratamento governamental, 1-2

Alzheimer, doença de, 1, 2, 3-4, 5

American Journal of Public Health, 1

Ames, Aldrich, 1-2

análise quantitativa, 1, 2

antibióticos, 1

aparelho de inferência bayesiana, 1-2, 3-4, 5, 6-7, 8

Application of Probability to Cryptography, The (Turing), 1

aquecimento global, 1, 2-3

Aristóteles, 1

Ars Conjectandi (Bernoulli), 1, 2

Atlantic City, 1, 2-3

Bacar, 1-2

Bachelier, Louis, 1

bancos de investimentos, 1, 2-3

Barnes, Steven, 1-2

Basic and Applied Social Psychology (Basp), 1-2, 3

Bayes, teorema de, 1-2

- DNA, perfil de, 1-2

- Great, estudo, 1, 2

- Innocence Project, 1-2

- ligaço com inferncia, 1-2

- problema dos a priori, 1

- rejeiço de Fisher do, 1-2

- usando o, 1-2, 3, 4-5, 6-7, 8

- uso de Turing do, 1-2

- verso de Turing do, 1-2

Bayes, Thomas:

- publicaço do trabalho, 1, 2

- regra de, 1-2, 3, 4-5

- trabalho de, 1-2, 3-4, 5

bayesianos, mtodos:

- controversos, 1-2, 3

correlação, 1
Great, estudo, 1-2
livros-texto, 1
termo, 1
uso em quebra de códigos, 1-2
usos, 1, 2, 3

Behind the Headlines (site), 1
Bernoulli, Daniel, 1-2, 3
Bernoulli, Jacob, 1, 2
Bernoulli, Nicolau, 1, 2
Betfair, 1, 2
Big Data, 1, 2-3, 4
Birmingham, Seis de, 1-2
blackjack (vinte e um), 1-2, 3, 4
Bletchley Park, 1-2, 3
BMJ, 1, 2
boi, adivinhar o peso do, 1-2
bolha da internet, 1, 2
Borel, Émile, 1
Breivik, Anders, 1
bruxa de Agnesi, curva, 1-2
Buffett, Warren, 1, 2

Cage, Nicolas, 1-2, 3
Call of Duty (videogame), 1-2

Calment, Jeanne, 1

campos eletromagnéticos, 1

câncer:

- cerebral, 1

- de colo do útero, 1

- de fígado, 1

- de mama, 1-2, 3, 4-5

- de pulmão, 1-2

- diagnóstico do, 1-2, 3, 4-5

- e aleatoriedade, 1

- e riscos, 1, 2-3, 4

- e tratamentos, 1

- no pâncreas, 1

- pesquisa do, 1

Cardano, Girolamo, 1-2

Carnegie, Andrew, 1

caso-controle, estudos, 1-2, 3

cassinos, 1-2

Cauchy, distribuição de, 1-2, 3

CDOs (Collateralised Debt Obligations, Obrigações de Débito Colateralizadas),
1-2, 3

Ceres, descoberta de, 1, 2

Cern, 1-2

Christensen, Eric, 1, 2

chuva, chance de, 1, 2-3

cibercriminosos, 1

ciência forense, 1-2

cigarros, fumar, 1-2

Citigroup, 1-2

coincidências:

- correlações e, 1, 2

- “espantosas”, 1-2, 3

- leis que governam as, 1-2

- predição de, 1, 2

- Titanic*, história do, 1-2

Colossus, computador, 1

comer carne, 1-2

contagem de cartas, 1-2

contexto:

- adolescentes e videogames, 1

- argumentos de pesquisas, 1-2

- Great, estudo, 1

- métodos bayesianos, 1-2, 3, 4

- primeira lei da ausência de leis, 1-2

- testes diagnósticos, 1-2

- valor do dinheiro, 1-2

coorte, estudos de, 1-2, 3

correlação:

- autocorrelação, 1, 2

- causalidade e, 1, 2

- coeficientes de, 1-2, 3, 4, 5

confiabilidade, questões de, 1-2

crenças a priori, 1

dados, 1

dados, garimpagem de (*data mining*), 1-2

dados, limpeza de, 1, 2-3

“descobertas”, 1-2

fatores de confusão (“confundimento”), 1-2

Google Flu Trends (Tendências de Gripe do Google), 1-2

sabedoria das multidões, 1, 2

significância, 1

teoria da, 1-2

teoria moderna do portfólio (TMP), 1-2

corridas de cavalo:

apostas, 1-2

chances de ganhar, 1

dicas, 1

estratégia de apostas, 1-2, 3-4

favoritos, 1

probabilidades, 1-2, 3, 4

sequência perdedora, 1

variáveis de resultados, 1-2

Couzin, Iain, 1

crença(s):

a priori, 1, 2-3, 4, 5, 6, 7, 8, 9

aleatoriedade, 1-2

atualização, 1, 2-3, 4, 5, 6, 7
coletivas, 1
e estratégias de investimentos, 1
em Deus, 1-2
estratégias de jogos e apostas, 1, 2, 3-4
evidência a favor ou contra, 1-2, 3, 4
níveis de, 1-2, 3-4, 5, 6, 7
probabilidades e, 1-2
regra de Bayes, 1-2, 3, 4-5
subjativa(s), 1
Turing, trabalho de, 1-2
crise financeira, 1, 2-3, 4-5, 6-7, 8, 9
curva do sino:
 crise financeira, 1-2
 distribuição de Cauchy, 1-2
 distribuição enviesada, 1-2
 distribuição normal, 1-2, 3, 4-5, 6
 distribuições de lei de potência, 1-2
 estatura humana, 1, 2-3, 4
 eventos raros, 1, 2-3
 fórmula, 1-2, 3
 Galton, trabalho de, 1
 Gauss, trabalho de, 1-2
 Greenspan, alerta de, 1
 história, 1-2

- lançamento de moeda, 1-2
- Laplace, trabalho de, 1-2
- média e mediana, 1-2
- performance de funcionários, 1-2
- picos dentados, 1, 2, 3
- Quetelet, trabalho de, 1-2
- salários, 1-2
- teorema do limite central, 1-2, 3, 4, 5-6, 7, 8-9
- TVE, comparação, 1-2, 3
- curva log-normal, 1

D'Alembert, Jean-Baptiste le Rond, 1, 2-3

dados:

- Big Data, 1, 2-3
- correlacionados, 1
- garimpagem (*data mining*), 1, 2-3, 4
- limpeza, 1, 2-3
- quantidade de, 1-2, 3

Darktrace, 1-2

Davis-Stober, Clinton, 1-2

De Haan, Laurens, 1, 2

De Moivre, Abraham, 1, 2-3

Deal, Mike, 1-2

decifração de códigos, 1-2, 3, 4

decisão, teoria da, 1, 2-3

estratégia de negócios, 1-2
mudar de casa, 1-2
produto químico, 1-2
derivativos, 1, 2-3, 4
detectores de mentiras, 1-2
Deus, crença em, 1-2
Diaconis, Persi, 1
Dimon, Jamie, 1, 2-3
distribuição generalizada de valores extremos (GVE), 1
distribuição normal, 1, 2, 3-4, 5-6, 7-8, 9, 10
distribuições viesadas, 1-2
DNA, perfil de, 1-2
Doll, Richard, 1
Durand, David, 1

Edwards, Andrew, 1
Eliano, Claudio, 1
Embrechts, Paul, 1
enchente de dados, 1-2
energia, conservação de, 1-2
Enigma, máquina, 1, 2
espantosa máquina de bobagens, 1-2, 3, 4-5, 6-7
Essay Towards Solving a Problem in the Doctrine of Chances (Bayes), 1, 2, 3
Estados Unidos:
Centros de Controle de Doenças (CDC – Centers for Disease Control), 1-2

mercados do Tesouro, 1

estatinas, 1-2

estendidas, garantias *ver* garantias estendidas

Estudos Clínicos Randomizados (ECRs), 1-2, 3-4, 5

Euler, Leonhard, 1

evidência:

- como dar sentido a, 1-2
- confessional, 1-2
- forense, 1-2

expectativa de vida, 1

experimentos com animais, 1-2

falácia do jogador, 1-2

Fama, Eugene, 1-2

febre de lei de potência, 1-2

Feynman, Richard, 1

Finetti, Bruno de, 1

Fisher, Ronald Aylmer, 1-2, 3, 4-5, 6-7, 8

frequências relativas:

- Bernoulli, trabalho de, 1-2
- Cardano, trabalho de, 1
- coincidências, 1-2
- comparação de, 1-2
- lei das médias, 1-2, 3, 4, 5, 6
- primeira lei da ausência de leis, 1-2, 3, 4-5

Rosto de Marte, 1
frequentismo, 1, 2-3, 4
fundos de hedge, 1-2
fundos de índice rastreador, 1-2
fundos de rastreador *ver* fundos de índice rastreador
futebol:
 jogos, 1, 2, 3, 4
 times, 1-2, 3

Galton, sir Francis, 1, 2-3, 4
garantias estendidas, 1, 2-3, 4
Gauss, Carl, 1-2, 3, 4
GCHQ, 1-2, 3
GEC-Marconi, 1
Gibbs, Josiah Willard, 1
Ginsburg, Norman, 1
Goldman Sachs, 1
Good, I.J. “Jack”, 1-2
Google Flu Trends (GFT), 1-2
Grande Colisor de Hádrons (LHC), 1, 2
grandes boladas de prêmio, 1-2, 3, 4, 5-6
Great, estudo, 1-2, 3-4
Greenspan, Alan, 1-2, 3, 4
Greiss, teste de, 1-2
grupos de controle, 1-2

Guarda Costeira dos Estados Unidos, 1-2, 3-4

Guildford, Quatro de, 1

Hamilton, Sue, 1-2

Hanlon, Michael, 1

Hertwig, Ralph, 1

Herzog, Stefan, 1

Hewlett-Packard, 1

Higgs, partícula de, 1-2, 3

Hill, Austin Bradford, 1, 2

HIV, diagnóstico, 1-2

Hollywood Stock Exchange (HSX), 1

Hong, Lu, 1

HPV, vacinação, 1-2

Hunt, Jeremy, 1

Hussain, Nasser, 1

incêndios florestais, 1, 2

independência:

- jogos de cartas, 1-2

- lançamento de moeda, 1

- loteria, 1-2

- premissa de, 1-2

Innocence Project, 1, 2-3

intervalos de confiança (ICs), 1

inveja da física, 1-2, 3

investimentos, 1, 2, 3, 4, 5-6

Ioannidis, John, 1

Iowa Electronic Market (Mercado Eletrônico de Iowa, IEM), 1

Jagger, Joseph, 1-2

Jeffreys-Lindley, paradoxo de, 1

jogos de azar, 1

Johnson, Donald, 1

Kahneman, Daniel, 1

Kao, Albert, 1

Kashiwagi, Akio, 1-2

Keillor, Garrison, 1

Kerrich, John, 1-2

Labouchère, sistema de, 1

Lake Wobegon, 1

lançamento de moeda:

- comportamento da moeda, 1

- curva do sino, 1-2

- D'Alembert, trabalho de, 1-2

- eventos casuais, 1, 2

- Kerrich, experimento de, 1-2

- lei das médias, 1, 2-3, 4-5

- maré de “azar”, 1

- predição, 1

Laplace, Pierre Simon de:

- curva do sino, trabalho da, 1-2, 3-4

- princípio da indiferença (princípio da razão insuficiente), 1

- problema dos a priori, 1

- teorema do limite central, 1-2, 3, 4, 5-6, 7, 8-9

Las Vegas, 1, 2-3, 4, 5-6, 7

lei das médias:

- Bernoulli, trabalho de, 1

- coleta de dados, 1

- estratégia de apostas, 1-2

- eventos casuais, 1

- jogos de cassino, 1-2, 3-4, 5-6

- lançamento de moeda, 1, 2-3, 4

- lei fraca dos grandes números, 1

- leis da ausência de leis, 1

- prêmios de seguros, 1

- problema da gaveta de meias, 1

- significado, 1, 2

lei de potência, distribuição de, 1-2, 3

lei dos grandes números, 1, 2, 3, 4, 5

lei fraca dos grandes números, 1, 2, 3, 4, 5

lei normal, 1

leis da ausência de leis:

- primeira, 1-2, 3-4, 5

- segunda, 1-2

terceira, 1

Less4U Ltda., 1-2

leucemia infantil, 1, 2

Lévy, voos de, 1, 2

Lévy-estáveis, distribuições, 1, 2

Lippmann, Gabriel, 1

Lo, Andrew, 1, 2, 3, 4

loterias:

aleatoriedade, 1-2

ganhar, 1-2

jogos de cassino, 1

primeira lei da ausência de leis, 1

Lynch, Peter, 1

Malkiel, Burton, 1

mamografia, 1-2

máquina de descobertas, 1

Markowitz, Harry, 1-2

mediana, 1-2

Menzies, William, 1

México, programa de bem-estar social, 1

Million Women Study, 1

Misco, Walter e Linda, 1-2, 3

Monte Carlo, 1

Mordin, Nick, 1

Morton, Natalie, 1-2

Mueller, Mark, 1, 2, 3, 4

Nasa, 1, 2

Nature, 1, 2, 3

Netflix, 1

Newton, Isaac, 1, 2, 3, 4, 5

Newton-John, Olivia, 1

Neyman, Jerzy, 1

observacionais, estudos, 1, 2-3

Orange Telecom, 1-2

Ortner, Gerhard, 1

ovos, com gema dupla, 1-2, 3

Page, Scott, 1

palpites “chutados”, 1, 2, 3, 4, 5, 6

paradoxo do aniversário, 1

pareidolia, 1

Pascal, Blaise, 1-2, 3-4, 5, 6, 7-8, 9, 10

Pearson, Karl, 1, 2, 3-4

pedras num jarro, 1-2

pênis, comprimento do, 1

Petrarca, 1-2

Pocock, Stuart, 1, 2, 3-4

Poincaré, Henri, 1

“poucos por cento”, regra dos, 1-2, 3

predição:

avaliação de filmes, 1

bayesiana, 1-2

Bernoulli, trabalho de, 1

clima, 1-2

coincidências, 1, 2

curva do sino, 1, 2, 3-4, 5, 6

deflagrador de Alzheimer, 1-2

distribuição generalizada de valores extremos (GVE), 1

Google Flu Trends, 1-2

Great, estudo, 1-2

lançamento de moeda, 1, 2

leis da probabilidade, 1

mercados, 1-2, 3

missões da Nasa, 1

números de loteria, 1-2, 3-4

palpites chutados, 1

planeta Ceres, 1

primeira lei da ausência de leis, 1

regressão linear, 1-2

roleta, 1

teoria dos valores extremos (TVE), 1-2

terremotos, 1-2

previsão do tempo, 1-2

Price, Richard, 1, 2, 3

Prince, Chuck, 1-2

princípio da precaução, 1

probabilidade:

- aleatória, 1, 2-3

- cassinos, 1, 2

- curva do sino, 1-2, 3

- diagnóstico de câncer de mama, 1-2

- epistêmica, 1, 2, 3

- estratégias de investimentos, 1-2

- existência de Deus, 1-2

- frequências relativas, 1-2

- frequentismo, 1-2

- leis da, 1-2, 3, 4, 5

- loterias, 1, 2

- mercados financeiros, 1-2, 3, 4

- probabilidades condicionais, 1, 2, 3-4, 5, 6, 7

- seguradoras, 1

- significado de, 1, 2-3

- teorema áureo, 1-2

- teorema de Bayes, 1-2, 3-4, 5, 6, 7-8

- teoria da, 1-2, 3-4, 5, 6, 7, 8, 9, 10

- tipos de, 1

- Turing, trabalho de, 1, 2, 3

- valor p , 1-2

problema dos a priori, 1, 2-3, 4, 5, 6, 7

prova matemática, 1

Przybylski, Andrew, 1

QI, resultados de testes de, 1-2, 3-4

Quetelet, Adolphe, 1-2, 3-4

Ramsey, Frank, 1

razão de probabilidade (RP), 1-2, 3-4, 5-6

regra áurea das apostas, 1-2, 3-4, 5

regressão:

- à média, 1-2

- com base em computador, 1

- linear, 1-2, 3-4

regressão linear *ver* regressão

Revell, Ashley, 1-2, 3

Richards, Donald, 1

risco, conceito de, 1

riscos para a saúde, 1-2

Ritz, cassino, 1

Robertson, Morgan, 1-2

Rogan, Bud, 1

roleta, 1-2, 3-4, 5-6, 7, 8-9

Roses, Allen, 1

Rosto de Marte, 1

Rothman, Kenneth, 1

Royal Society, 1-2, 3

Rumsfeld, Donald, 1

sabedoria das multidões, 1-2, 3

Samuelson, Paul, 1-2

São Petersburgo, paradoxo de, 1-2, 3

Sarops (Search and Rescue Optimal Planning System), 1-2

Science, 1, 2

Segal, Tom, 1

seguros, 1, 2-3, 4-5, 6-7

Shannon, Claude, 1

Siemens, 1

significância estatística, 1-2, 3-4, 5, 6-7, 8, 9

Spiegelhalter, David, 1, 2, 3-4

Statistical Methods for Research Workers (Fisher), 1, 2

suicídios, 1-2, 3-4

Swensen, David, 1

Székely, Gabor, 1

Tell, Guilherme, 1

teorema áureo, 1-2

teorema do limite central, 1-2, 3, 4, 5-6, 7, 8-9

teoria dos valores extremos (TVE), 1, 2-3

teoria moderna do portfólio (TMP), 1-2

terremotos:

 predição, 1-2

registros, 1-2

teste de significância:

técnica falha, 1-2, 3-4, 5, 6

uso do, 1-2, 3, 4, 5-6, 7-8

Thibodeaux, Damon, 1-2, 3

Thorp, Ed, 1, 2-3

Tippett, Leonard, 1

Titanic, desastre do, 1-2

Triângulo das Bermudas, 1

Turing, Alan, 1-2, 3, 4-5

Tversky, Amos, 1

UrEDAS (Sistema Urgente de Detecção e Alarme de Terremotos), 1

valor médio (a “média”), 1-2

valor p , 1-2, 3-4, 5-6, 7, 8, 9-10

Veitch, Patrick, 1-2, 3, 4, 5

Venona, Projeto, 1-2

Vigen, Tyler, 1-2

Viniar, David, 1-2, 3-4, 5

Wadlow, Robert, 1

Which?, revista, 1-2

Wilmott, Paul, 1, 2, 3

Wilson, Edwin, 1

Winfield, John, 1

Worcester, sir Robert, 1

Título original:

Chancing It

(The Laws of Chance and How They Can Work for You)

Tradução autorizada da primeira edição inglesa, publicada em 2016 por Profile Books Ltd., de Londres, Inglaterra

Copyright © 2016, Robert Matthews

Copyright da edição brasileira © 2017:

Jorge Zahar Editor Ltda.

rua Marquês de S. Vicente 99 – 1º | 22451-041 Rio de Janeiro, RJ

tel (21) 2529-4750 | fax (21) 2529-4787

editora@zahar.com.br | www.zahar.com.br

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação de direitos autorais.
(Lei 9.610/98)

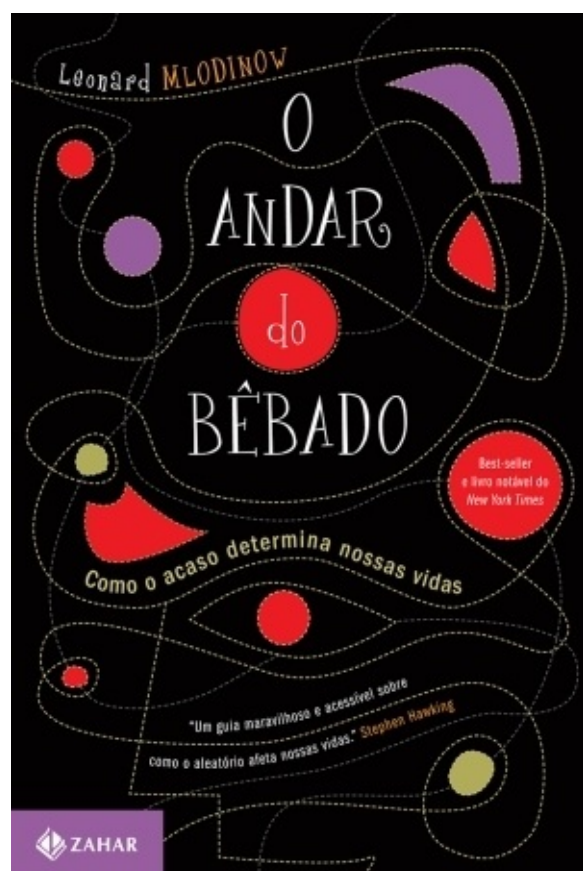
Grafia atualizada respeitando o novo Acordo Ortográfico da Língua Portuguesa

Capa: Estúdio Insólito | Imagem da capa: © Robin Atkins/Getty Images

Produção do arquivo ePub: Booknando Livros

Edição digital: abril de 2017

ISBN: 978-85-378-1669-1



O andar do bêbado

Mlodinow, Leonard

9788537801819

322 páginas

[Compre agora e leia](#)

Best-seller internacional e livro notável do New York Times

Um dos 10 Melhores Livros de Ciência, segundo a Amazon.com

Não estamos preparados para lidar com o aleatório e, por isso, não percebemos o quanto o acaso interfere em nossas vidas. Num tom irreverente, citando exemplos e pesquisas presentes em todos os âmbitos da vida, do mercado financeiro aos esportes, de Hollywood à medicina, Leonard Mlodinow apresenta de forma divertida e curiosa as ferramentas necessárias para identificar os indícios do acaso. Como resultado, nos ajuda a fazer escolhas mais acertadas e a conviver melhor com fatores que não podemos controlar.

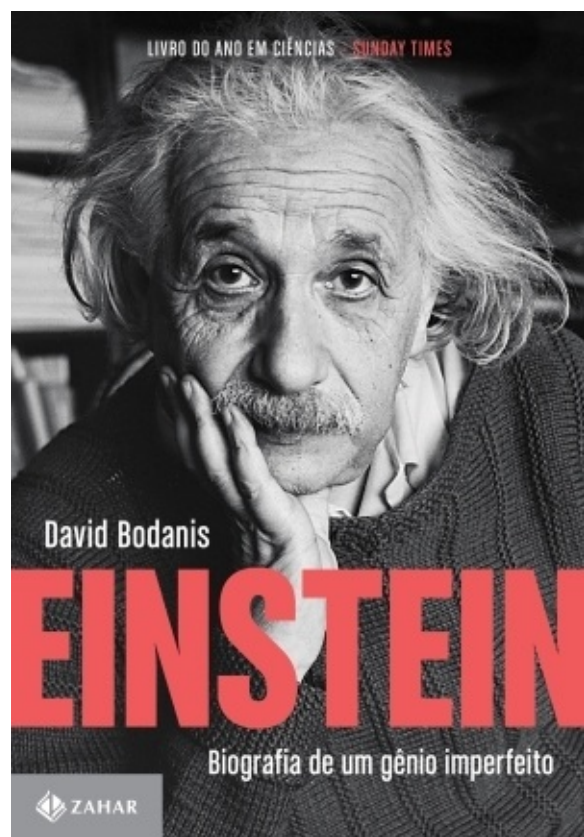
Prepare-se para colocar em xeque algumas certezas sobre o funcionamento do mundo e para perceber que muitas coisas são tão previsíveis quanto o próximo passo de um bêbado depois de uma noitada...

"Um guia maravilhoso e acessível sobre como o aleatório afeta nossas vidas" Stephen Hawking

"Mlodinow escreve num estilo leve, intercalando desafios probabilísticos com perfis de cientistas... O resultado é um curso intensivo, de leitura

agradável, sobre aleatoriedade e estatística." George Johnson, New York Times

[Compre agora e leia](#)



Einstein

Bodanis, David
9788537816714
282 páginas

[Compre agora e leia](#)

Um mergulho irresistível no lado mais humano de Einstein

Acessível e absorvente, esta não é apenas mais uma biografia do maior gênio da era moderna. Escrita pelo premiado autor David Bodanis, retrata o cientista revolucionário para revelar um Albert Einstein profundamente humano em sua genialidade e em seus defeitos e imperfeições – entre eles a teimosia orgulhosa que o deixou isolado e à margem da comunidade científica nas últimas décadas de vida.

A chegada de um gênio ao ápice, seu declínio, o modo como lidou com o fracasso e a perda da confiança – esse é o mapa percorrido por Bodanis nessa reconstrução minuciosa e afetiva, mas também crítica, da vida de Einstein.

Com uma narrativa cativante, o livro oferece ainda explicações científicas ao alcance do leitor não especializado – que ficará surpreso ao descobrir que é possível entender a teoria da relatividade geral.

"Arrebatador!" Forbes

"Ninguém torna a ciência complexa mais fascinante e acessível que

David Bodanis." Bill Bryson

"Sensível e perspicaz, mostra o modo como um gênio pode perder a majestade dentro da comunidade científica." Sunday Times

"Um grande prazer. Bodanis dá voz às mulheres na vida de Einstein, não faz julgamentos sobre o biografado e oferece um livro absolutamente envolvente e revelador." Shelf Awareness

"Uma biografia que admira seu personagem, mas não se furta a criticá-lo, mostrando o grande gênio tomado pelo pensamento inflexível em seus últimos anos." The Observer

"Habilidade extraordinária para explicar as questões mais complicadas ... Teorias do universo se transformam em teorias da vida." The Times

[Compre agora e leia](#)

Inclui posfácio do autor sobre o Brasil

REDES Manuel Castells DE INDIGNAÇÃO E ESPERANÇA



Movimentos sociais
na era da internet



Redes de indignação e esperança

Castells, Manuel

9788537811153

272 páginas

[Compre agora e leia](#)

Principal pensador das sociedades conectadas em rede, Manuel Castells examina os movimentos sociais que eclodiram em 2011 - como a Primavera Árabe, os Indignados na Espanha, os movimentos Occupy nos Estados Unidos - e oferece uma análise pioneira de suas características sociais inovadoras: conexão e comunicação horizontais; ocupação do espaço público urbano; criação de tempo e de espaço próprios; ausência de lideranças e de programas; aspecto ao mesmo tempo local e global. Tudo isso, observa o autor, propiciado pelo modelo da internet.

O sociólogo espanhol faz um relato dos eventos-chave dos movimentos e divulga informações importantes sobre o contexto específico das lutas. Mapeando as atividades e práticas das diversas rebeliões, Castells sugere duas questões fundamentais: o que detonou as mobilizações de massa de 2011 pelo mundo? Como compreender essas novas formas de ação e participação política? Para ele, a resposta é simples: os movimentos começaram na internet e se disseminaram por contágio, via comunicação sem fio, mídias móveis e troca viral de imagens e conteúdos. Segundo ele, a internet criou um "espaço de autonomia" para a troca de informações e para a partilha de sentimentos coletivos de indignação e esperança - um novo modelo de participação cidadã.

[Compre agora e leia](#)

J.-D. Nasio

9 Lições sobre
Arte e Psicanálise



 ZAHAR

9 lições sobre arte e psicanálise

Nasio, J.-D.

9788537816707

142 páginas

[Compre agora e leia](#)

A ótica particular do renomado psicanalista argentino em análises sobre pintura, música e dança

Nas palavras do psicanalista e psiquiatra J.-D. Nasio, a concepção de uma obra de arte "é um processo único e impenetrável. É impossível surpreender o segredo do ato de criar, que permanecerá sempre um mistério. A única coisa que podemos fazer é reconstruir mentalmente, a posteriori, o momento criador, e mesmo isso só é possível até certo ponto. Uma vez que nos é proibido partilhar com o artista seu ato de criar, não nos resta senão tentar revivê-lo em imaginação, senti-lo."

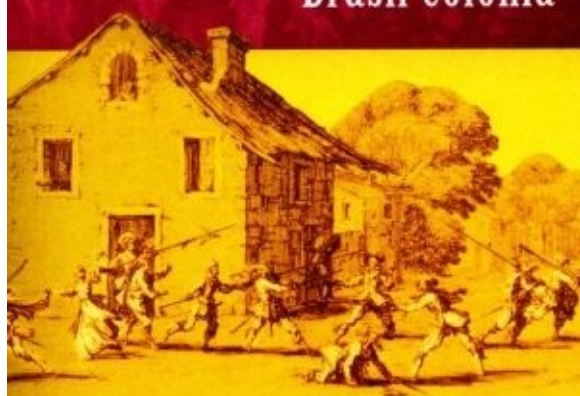
Partindo dessa premissa, Nasio investiga a questão da criação em vários âmbitos da arte. Na música, explora a singularidade da experiência de Maria Callas, para muitos a diva máxima do canto lírico. No campo da pintura, investiga a relação entre as obras de Francis Bacon e Velázquez; analisa o impacto que uma tela de Picasso teve sobre uma paciente; e se debruça sobre os expressivos quadros de Félix Vallotton, numa avaliação profunda das experiências pessoais do pintor e da influência delas sobre sua vasta produção. E não deixa de lado a dança, como uma sublime expressão do inconsciente.

Em jogo ao longo de todo o livro estão a arte e a sublimação - conceitos que Nasio relaciona de forma original e ao mesmo tempo profundamente ligada ao pensamento psicanalítico.

[Compre agora e leia](#)

JORGE ZAHAR EDITOR

Rebeliões no Brasil Colônia



LUCIANO FIGUEIREDO

Descobrindo o Brasil

Rebeliões no Brasil Colônia

Figueiredo, Luciano

9788537807644

88 páginas

[Compre agora e leia](#)

Inúmeras rebeliões e movimentos armados coletivos sacudiram a América portuguesa nos séculos XVII e XVIII. Esse livro propõe uma revisão das leituras tradicionais sobre o tema, mostrando como as lutas por direitos políticos, sociais e econômicos fizeram emergir uma nova identidade colonial.

[Compre agora e leia](#)